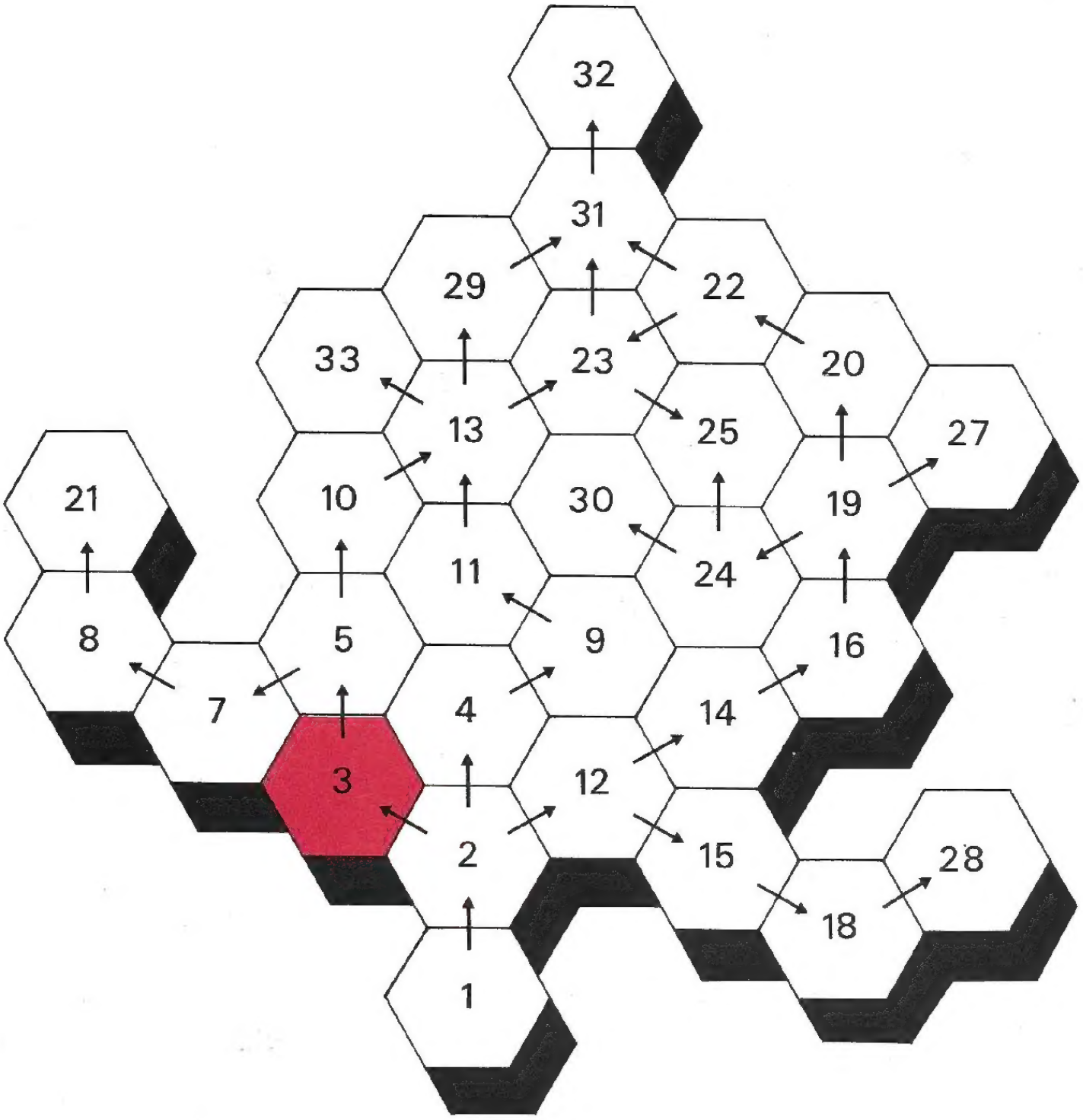
M201·3

THE OPEN UNIVERSITY

Mathematics: A Second Level Course

Linear Mathematics Unit 3

Hermite Normal Form



The Open University

Mathematics: A Second Level Course

Linear Mathematics Unit 3

HERMITE NORMAL FORM

Prepared by the Linear Mathematics Course Team

The Open University Press

The Open University Press Walton Hall MK7 6AA

First published 1972. Reprinted 1976

Designed by the Media Development Group of the Open University.

Printed in Great Britain by
Martin Cadbury

SBN 335 01092 X

This text forms part of the correspondence element of an Open University Second Level Course. The complete list of units in the course is given at the end of this text.

For general availability of supporting material referred to in this text, please write to the Director of Marketing, The Open University, P.O. Box 81, Walton Hall, Milton Keynes, MK7 6AT.

Further information on Open University courses may be obtained from the Admissions Office, The Open University, P.O. Box 48, Walton Hall, Milton Keynes, MK7 6AB.

Contents

Set Books

D. L. Kreider, R. G. Kuller, D. R. Ostberg and F. W. Perkins, *An Introduction to Linear Analysis* (Addison-Wesley, 1966).

E. D. Nering, *Linear Algebra and Matrix Theory* (John Wiley, 1970).

It is essential to have these books; the course is based on them and will not make sense without them.

Conventions

Before working through this correspondence text make sure you have read *A Guide to the Linear Mathematics Course*. Of the typographical conventions given in the Guide the following are the most important.

The set books are referred to as:

K for *An Introduction to Linear Analysis*

N for *Linear Algebra and Matrix Theory*

All starred items in the summaries are examinable.

References to the Open University Mathematics Foundation Course Units (The Open University Press, 1971) take the form *Unit M 100 3, Operations and Morphisms.*

3.0 INTRODUCTION

This unit is more down-to-earth than the last two. It is concerned with two main topics. First (in Section 3.1), it looks at the practical details of how vectors and linear transformations are represented in numerical terms by matrices, and how to calculate the effect on these matrices of changing the bases. Second (in the rest of the unit), it builds up a systematic technique for calculating the solutions of *linear problems*. (By a linear problem, we mean one of the form

"find all ξ in V such that $\sigma(\xi) = \alpha$"

where σ is a linear transformation of a vector space V and α is a given vector in its codomain.)

Before we can even express a linear problem satisfactorily, we need to be able to bring our vectors and transformations down to earth, by selecting bases of the vector spaces and expressing the vectors and transformations in terms of matrices with respect to these bases. You saw how this could be done in *Unit 2, Linear Transformations*. That unit did not, however, discuss the problem of deciding *which bases to use*. The matrices are likely to be simpler in form for some bases than for others, and we would like to know how to select the "best" bases with respect to which the problem can be expressed.

This means that we are likely to want to change bases fairly frequently, and it would be desirable to know what happens to a matrix representing a linear transformation when we do change bases. Also, can we describe basis changes themselves in terms of matrices?

In Section 3.1, we obtain satisfactory answers to all these questions. If U is an n-dimensional vector space, then we can describe a change from an "old" basis $\{\alpha_1, \ldots, \alpha_n\}$ to a "new" basis $\{\alpha'_1, \ldots, \alpha'_n\}$ in terms of an $n \times n$ *transition matrix* P, whose columns are just the components of the new basis vectors in terms of the old. Then if X is a one-column matrix describing the components of a fixed vector ξ in terms of the "old" basis, while X' describes ξ in terms of the "new" basis, we find that X and X' are related by the matrix equation

$$X = PX' \tag{1}$$

We also find matrix equations to describe what happens to the matrices of linear transformations. Suppose $\sigma: U \longrightarrow V$ is a linear transformation, and its matrix with respect to "old" bases in U and V is A. Then we find that if we change the basis in U by a transition matrix P, we get a new matrix for σ, namely

$$A' = AP. \tag{2}$$

On the other hand, if we change the basis in V by a transition matrix Q, we get, for the new matrix for σ:

$$A'' = Q^{-1}A \tag{3}$$

and if we change *both* bases at once, we get the matrix

$$A''' = Q^{-1}AP. \tag{4}$$

Furthermore, if we allow ourselves the liberty to change both bases in this way, then we can get a very simple form for A''', namely

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdot & \cdots & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & \cdot & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

in which the "diagonal" entries $a_{11}, a_{22}, \ldots, a_{\rho\rho}$ are 1s (where ρ is the rank of the transformation), and *all* other elements are 0s.

However, for the purpose of solving linear problems, it turns out that the appropriate thing to do is to change the basis of the *codomain only*, i.e. to multiply the matrix of the problem *on the left* by another (non-singular) matrix. To see this, let us look at linear problems in more detail.

The problem of solving a system of linear equations such as

$$\begin{aligned} 2x_1 - x_2 &= 4 \\ x_1 - x_2 &= 5 \end{aligned}$$

is a linear problem; so is the problem of solving the differential equation

$$X''(t) + X(t) = \cos pt \qquad (t \in R).$$

In the present unit we shall be concerned only with the case where the domain and codomain have finite dimension, so that σ has a matrix representation; that is, with problems like the first of the above examples, which can be written in the matrix form

$$AX = Y,$$

rather than the second, which cannot. If the matrix A is non-singular and we know A^{-1}, this equation can be solved by pre-multiplying (i.e. multiplying on the left) by A^{-1}, to obtain the solution in the form

$$X = A^{-1}Y.$$

For example, in the system of equations above we have

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad A = \begin{bmatrix} 2 & -1 \\ 1 & -1 \end{bmatrix}, \qquad Y = \begin{bmatrix} 4 \\ 5 \end{bmatrix}.$$

Thus $A^{-1} = \begin{bmatrix} 1 & -1 \\ 1 & -2 \end{bmatrix}$, and the solution is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \end{bmatrix} = \begin{bmatrix} -1 \\ -6 \end{bmatrix}.$$

In practice, we may not know A^{-1}, but the same result is achieved by manipulating the equations by a succession of elementary operations (such as adding a multiple of one equation to another), whose net effect is virtually the same as multiplying on the left by A^{-1}. This is the Gauss elimination method, described in *Unit M100 26, Linear Algebra III.*

A difficulty which we touched upon in the Foundation Course is that A may not have an inverse, as in the following system of equations in which there are more variables than equations.

$$\begin{aligned} 2x_1 - x_2 + x_3 &= 4 \\ x_1 - x_2 - x_3 &= 5. \end{aligned}$$

For such equations, no amount of manipulation will give us a unique solution; either there is no solution at all, or there is an infinite set of solutions. (In the above example, the solution set is infinite.)

What we would like, therefore, is a generalization of the Gauss elimination method which will deal with *any* system of equations, whether its matrix has an inverse or not; and which will in all cases lead us to the solution set, whether it consists of a unique solution, is empty, or is infinite.

The method is just like the Gauss elimination method, in that we manipulate the equations by elementary operations. The net effect of these operations (as we shall see) is to multiply the original equation $AX = Y$ on the left by a matrix which may or may not be A^{-1}. We call it Q^{-1} instead; thus the final form of the equations is

$$(Q^{-1}A)X = Q^{-1}Y,$$

and we choose Q^{-1} to make the matrix $Q^{-1}A$ as "simple" as we can, if possible, the unit matrix. This simplest possible matrix is called the *Hermite normal form* of the matrix A and is the main object of study in this unit. The usefulness of Hermite normal form is not confined to the solution of linear equations: it can also be used to find the inverse mapping of any linear transformation, to find the linear relations that exist among any set of vectors, and to find a basis for a subspace of a vector space.

Since going from the matrix A to the matrix $Q^{-1}A$ is equivalent to changing the codomain basis by a transition matrix Q, it follows that the problem of defining Hermite normal form turns out to be equivalent to the problem of choosing the most convenient basis in the codomain. In the second section of this unit we show how to define such a basis, and use it to specify the Hermite normal form completely. We also use Hermite normal form to prove that the row and column ranks of a matrix are equal.

In the later sections of the unit we show how to calculate Hermite normal forms and consider various applications: techniques for inverting matrices, solving linear problems, and finding linear relations among sets of vectors.

A computer program for finding the Hermite normal form of a matrix and its inverse is provided in the supplementary material. It may be used for some of the exercises in Section 3.3.

3.1 CHANGE OF BASIS

3.1.0 Introduction

This section contains the theory underlying a great deal of the manipulations with vectors and matrices that we shall do subsequently. This material is central to what follows in this and subsequent units; consequently, it is important. It is probable that you may find this section confusing at first, but we expect that after you have used this material it will seem almost obvious to you.

We know that there are certain statements in linear algebra which *do not* depend on the choice of basis, such as whether a linear transformation is an isomorphism; but, there are also statements which *do* depend on the choice of basis, such as what the entries are in the matrix representing a linear transformation. In this section we examine what happens to statements of this second kind, when the basis is changed. Once we know this, we can change bases to our advantage in calculations, because in certain bases, n-tuples and matrices may take particularly simple forms. For example, the law of matrix multiplication is particularly easy to use when the matrix is diagonal:

$$\text{the inverse of } \begin{bmatrix} 7 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 8 \end{bmatrix} \quad \text{is} \quad \begin{bmatrix} \frac{1}{7} & 0 & 0 \\ 0 & \frac{1}{12} & 0 \\ 0 & 0 & \frac{1}{8} \end{bmatrix}$$

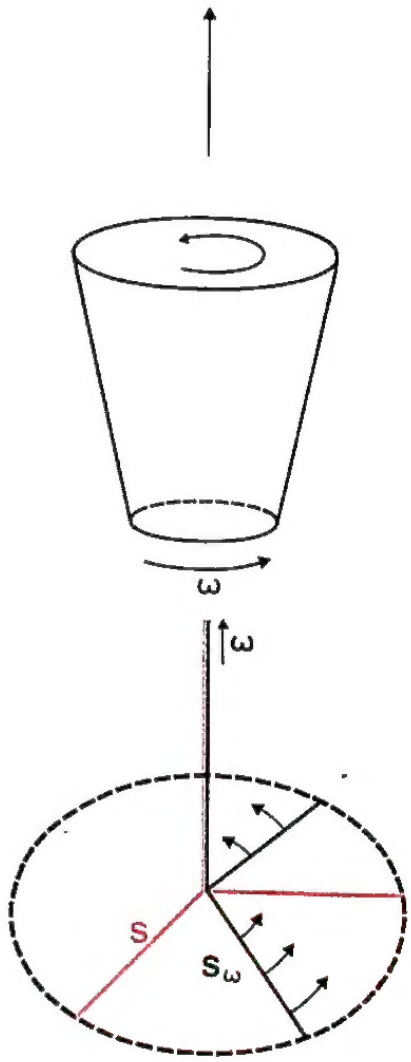
It will therefore be useful (if possible) to be able to convert the matrix representing a linear transformation with respect to one basis into a diagonal matrix representing the same linear transformation with respect to some other basis.

Another example, often used in scientific applications, is the rotation of Cartesian coordinate systems. It is often useful (and sometimes necessary), to be able to describe the same vector in different coordinate systems. For example, suppose that we were trying to describe the bulk motion of a fluid, enclosed in a bucket which is rotating at a constant angular frequency, ω.

Let S be the coordinate system of an observer at rest who will see the fluid moving around the bucket, with a combination of the bucket's rotatory motion and various other types of motion (waves, etc.). Let S_ω be a coordinate system which is rotating with the bucket.

An observer attached to the rotating system S_ω would see a very simple state of motion of the fluid therefore, and could readily "do the physics" for the problem. But to find the solution to the original problem, our S_ω-observer's results must be transformed to results valid in S. Thus, we must know how vectors transform from S to S_ω. This is just a rotation of angle $\theta = \omega t$ (t is time), and is a *change of basis* of a special sort.

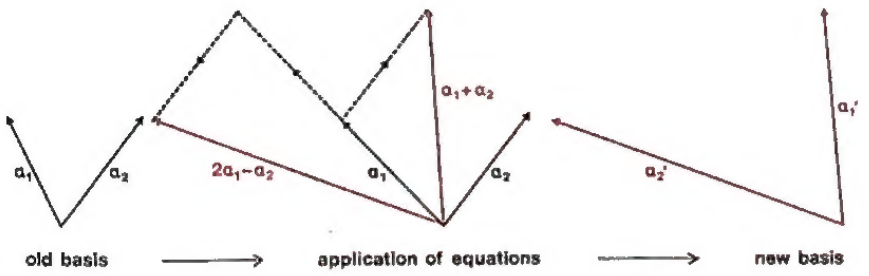
The key to the technique of changing bases is a matrix, called the *matrix of transition* for the basis change being considered; we shall find that its columns are the coordinates of the *new basis vectors* with respect to the *old basis*. Using the appropriate matrix of transition, we can go from the coordinates of any given vector with respect to one basis to its coordinates with respect to another by a single matrix multiplication. In the case of linear transformations the situation is a little more complicated, because we can change the bases in the domain and in the codomain at the same time, but these basis changes can still be carried out very conveniently using the appropriate matrices of transition. In fact, these changes may be viewed as arising out of two matrices of transition applied simultaneously.



3.1.1 The Matrix of Transition

As an example to begin with, suppose that the vector space U is two-dimensional. Call $\{\alpha_1, \alpha_2\}$ the old basis for U and suppose that new basis vectors for U are α'_1, α'_2 given by

$$\alpha'_1 = \alpha_1 + \alpha_2$$
$$\alpha'_2 = 2\alpha_1 - \alpha_2$$



Note that $\{\alpha'_1, \alpha'_2\}$ does indeed constitute a basis for U if $\{\alpha_1, \alpha_2\}$ does. As we have seen in the previous unit (***Theorem 1.17***, page N34), there exists a unique linear transformation that maps α_1 to α'_1 and α_2 to α'_2. For ***Theorem 1.17*** tells us that if $A = \{\alpha_1, \alpha_2\}$ and $B = \{\alpha'_1, \alpha'_2\}$, then:

$$\begin{aligned} \sigma(\alpha_1) &= \alpha'_1 = \alpha_1 + \alpha_2 \\ \sigma(\alpha_2) &= \alpha'_2 = 2\alpha_1 - \alpha_2 \end{aligned} \tag{1}$$

completely defines the (change of basis) linear transformation σ. This transformation can be represented by a matrix in the usual way, with respect to any basis in the domain and any basis in the codomain. Since the domain and codomain, in this case, are the same space U, it is natural to use the same basis for both; if we take this basis to be $\{\alpha_1, \alpha_2\}$ then, as we saw in the previous unit, the columns of the matrix are the coordinates of the images of α_1 and α_2 with respect to this basis. Applying Equation (2.2) on page N38, with $A = B = \{\alpha_1, \alpha_2\}$, we have:

$$\begin{aligned} \sigma(\alpha_1) &= a_{11}\alpha_1 + a_{21}\alpha_2 \\ \sigma(\alpha_2) &= a_{12}\alpha_1 + a_{22}\alpha_2 \end{aligned} \tag{2}$$

By comparing Equations (1) and (2), we see that the matrix representing the transformation from the new basis to the old, *with respect to the old basis*, is

$$\begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}.$$

This is called the *matrix of transition** for this particular change of basis. To single out its particular use, we use the symbol P for a transition matrix, and p_{ij} for its entries. If there are two transition matrices being used in the same problem, we shall use the symbols P and Q.

The corresponding formula for a general change of basis is given in the next reading passage.

READ from the top of page N50 *to the sentence* "Thus P is non-singular." *about half-way down the page.*

Note

Just as in Equation (2.2) on page N38, so in Equation (4.1) on page N50, we sum over the first of the double suffices, the one labelling the rows. This is the way it is always done in an equation involving matrices and vectors.

* We see how this matrix relates the coordinates of a vector with respect to the new and old bases shortly.

Exercises

1. Let $U = R^3$, $\alpha_1 = (1, 0, 0)$, $\alpha_2 = (0, 1, 0)$, $\alpha_3 = (0, 0, 1)$, $\alpha'_1 = (1, 1, 1)$, $\alpha'_2 = (2, 1, 1)$ and $\alpha'_3 = (1, -1, 0)$. Find the matrix of transition with respect to the old basis $\{\alpha_1, \alpha_2, \alpha_3\}$.
2. Exercise 1, page N52.

Solutions

1. We must start by finding out just what particular linear combination of $\{\alpha_1, \alpha_2, \alpha_3\}$ each of the α'_j is. Direct inspection shows us that

$$\alpha'_1 = \alpha_1 + \alpha_2 + \alpha_3$$
$$\alpha'_2 = 2\alpha_1 + \alpha_2 + \alpha_3$$
$$\alpha'_3 = \alpha_1 - \alpha_2 .$$

Now

$$\alpha'_1 = \sigma(\alpha_1) = p_{11}\alpha_1 + p_{21}\alpha_2 + p_{31}\alpha_3 , \text{ etc. gives}$$

$$P = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 0 \end{bmatrix}.$$

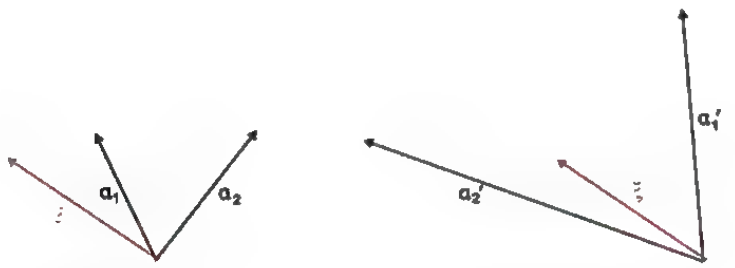
2. Writing $\alpha_1 = 1$, $\alpha_2 = x$, $\alpha_3 = x^2$, the new basis is given in terms of the old by

$$\alpha'_1 = p_1(x) = x^2 + x + 1 = \alpha_1 + \alpha_2 + \alpha_3$$
$$\alpha'_2 = p_2(x) = x^2 - x - 2 = -2\alpha_1 - \alpha_2 + \alpha_3$$
$$\alpha'_3 = p_3(x) = x^2 + x - 1 = -\alpha_1 + \alpha_2 + \alpha_3$$

and so the matrix of transition is $\begin{bmatrix} 1 & -2 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

3.1.2 Change of Basis for Coordinates of Vectors

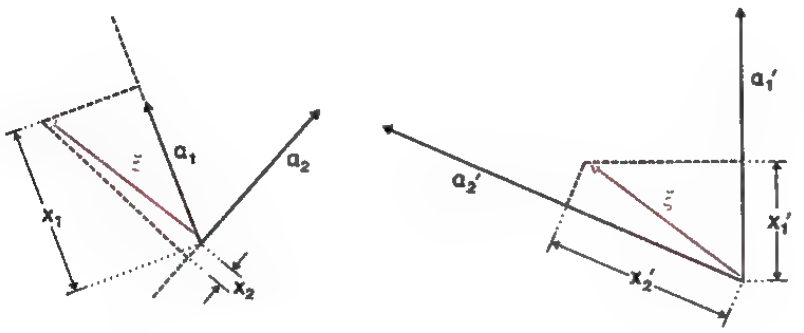
As our first application of the matrix of transition we consider the problem of finding the coordinates of vectors with respect to one basis when the coordinates with respect to another basis are given. Loosely speaking, since a vector can be written as $a_1\alpha_1 + a_2\alpha_2 + \cdots$, where the a_i are the coordinates and the α_j the basis vectors, if the basis vectors change in a certain way, the coordinates must change in a related way in such a fashion that *the vector itself does not change*. This is the crux of the matter. For, the vector exists as an entity in its own right (think of a geometric vector in space); the basis vectors are chosen independently of the existence of the vector in question. Let us use our example from sub-section 3.1.1 to illustrate the problem: we have calculated already the matrix of transition. Thus, $\{\alpha_1, \alpha_2\}$ is the old basis, $\{\alpha_1' = \alpha_1 + \alpha_2, \alpha_2' = 2\alpha_1 - \alpha_2\}$ is the new basis of U, and $P = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}$ is the relevant matrix of transition. Let ξ be a vector in U:



old basis; vector ξ — new basis; same vector ξ

Observe that ξ is the same in both cases. Once this point is grasped, the rest is arithmetic.

The coordinates of ξ, as opposed to ξ itself, do depend upon the basis chosen: they measure the "geometrical projection" of ξ on to the basis vectors.



Now

$$\xi = x_1\alpha_1 + x_2\alpha_2; \tag{1a}$$

but as well,

$$\xi = x_1'\alpha_1' + x_2'\alpha_2'. \tag{1b}$$

Because the α_j' are (known) linear combinations of $\{\alpha_1, \alpha_2\}$, the x_j' are related linear combinations (to be found) of $\{x_1, x_2\}$. Setting $\xi = \xi$ with the

left-hand side obtained from Equation (1a) and the right-hand side obtained from Equation (1b), we have

$$\begin{aligned} x_1\alpha_1 + x_2\alpha_2 &= x'_1\alpha'_1 + x'_2\alpha'_2 \\ &= x'_1(\alpha_1 + \alpha_2) + x'_2(2\alpha_1 - \alpha_2) \\ &= (x'_1 + 2x'_2)\alpha_1 + (x'_1 - x'_2)\alpha_2 . \end{aligned}$$

Linear independence of α_1 and α_2 now gives us

$$\begin{aligned} x_1 &= x'_1 + 2x'_2 \\ x_2 &= x'_1 - x'_2 \end{aligned} \tag{2}$$

These equations can be written in matrix form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix}$$

or

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = P \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix} \tag{2a}$$

where $P = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}$ is just the matrix of transition for this problem.

Notice that the equation gives the *old* coordinates in terms of the *new* ones, whereas in the preceding section we expressed the *new* basis vectors in terms of the *old* ones. The two concepts are "complementary". If we have the new basis vectors in terms of the old, then the natural substitution in any expression, is to replace the new basis vectors by the old. We then inevitably get the old coordinates in terms of the new ones. Look at the calculation we performed above, and make sure you see why this happens before reading the next piece of N. Note that Equations (2) or (2a) are a system of linear equations relating the coordinates, and that the matrix of coefficients is P. That is, we can write Equation (2a) in indexed form as

$$x_i = \sum_j p_{ij} x'_j .$$

Here the summation is over the *second* index not the first as in sub-section 3.1.1. This is another facet of the complementary nature of basis vectors and coordinates.

READ the rest of page N50.

Exercise

Use Equation (4.3) on page N50 to express the vector

$$\xi = -2\alpha'_1 + \alpha'_2 - 3\alpha'_3 \text{ in } P_3 ,$$

in terms of $\{\alpha_1 = 1, \alpha_2 = x, \alpha_3 = x^2\}$, where

$$\alpha'_1 = x^2 + x + 1, \alpha'_2 = x^2 - x - 2, \alpha'_3 = x^2 + x - 1.$$

Check by direct manipulation of the polynomials.

Solution

Here we have $X' = \begin{bmatrix} -2 \\ 1 \\ -3 \end{bmatrix}$ and $P = \begin{bmatrix} 1 & -2 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ as in the previous solution. Hence Equation (4.3) gives

$$X = \begin{bmatrix} 1 & -2 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} -1 \\ -6 \\ -4 \end{bmatrix}$$

by matrix multiplication.

That is, $\xi = -\alpha_1 - 6\alpha_2 - 4\alpha_3$.

Check $\xi = -2(x^2 + x + 1) + (x^2 - x - 2) - 3(x^2 + x - 1)$
$= -4x^2 - 6x - 1$
$= -\alpha_1 - 6\alpha_2 - 4\alpha_3$

as before.

3.1.3 Change of Basis in the Domain of a Linear Transformation

If σ is a linear transformation with domain U and codomain V, the matrix representing σ will change if we change the basis in either the domain or the codomain or in both. This is a matter of definition. The matrix entries a_{ij} in Equation (2.2) on page N38

$$\sigma(\alpha_j) = \sum_{i=1}^{m} a_{ij}\beta_i$$

clearly depend upon the basis vectors $\{\alpha_j\}$, $\{\beta_i\}$.

It is less complicated to consider a basis change in the domain only first, then a basis change in the codomain only, and finally put them together to form the general case.

The result will be concisely stated in matrix form, but deriving the result involves a certain amount of index manipulation. To help with this difficulty, we have devised an example. Therefore let us consider the case where U is 2-dimensional, with a basis $A = \{\alpha_1, \alpha_2\}$ and V is 3-dimensional, with a basis $B = \{\beta_1, \beta_2\ \beta_3\}$. We know from an earlier section (***Theorem 1.17***, page N34) that a linear transformation is completely determined by what it does to a basis in the domain. A general linear transformation $\sigma: U \longrightarrow V$ can therefore be specified by giving the images of α_1 and α_2. Suppose these are

$$\begin{aligned} \sigma(\alpha_1) &= r_1\beta_1 + r_2\beta_2 + r_3\beta_3 \\ \sigma(\alpha_2) &= s_1\beta_1 + s_2\beta_2 + s_3\beta_3 \end{aligned} \tag{1}$$

where $r_1, r_2, r_3, s_1, s_2, s_3$ are arbitrary scalars.

Let M be the matrix corresponding to σ with bases A in U, and B in V; then the columns of M are the coordinates of $\sigma(\alpha_1)$ and $\sigma(\alpha_2)$ with respect to the basis B; so Equation (1) gives

$$M = \begin{bmatrix} r_1 & s_1 \\ r_2 & s_2 \\ r_3 & s_3 \end{bmatrix}.$$

We already have a particular example of a basis change in U worked out in the preceding two sub-sections, where we change the basis in U from $\{\alpha_1, \alpha_2\}$ to $\{\alpha'_1, \alpha'_2\}$ with

$$\begin{aligned} \alpha'_1 &= \alpha_1 + \alpha_2 \\ \alpha'_2 &= 2\alpha_1 - \alpha_2 \end{aligned} \tag{2}$$

The matrix of transition was

$$P = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}. \tag{3}$$

Now $\{\alpha'_1, \alpha'_2\}$ is also a basis for U, and we can also find the matrix, M' representing the same linear transformation, σ, with respect to this *new* basis for the domain, $\{\alpha'_1, \alpha'_2\}$, and the *same* basis for the codomain $\{\beta_1, \beta_2, \beta_3\}$; this is what this sub-section is all about. First note that σ can be thought of as having an existence independent of bases, just as a vector does. It is only the representing matrix which depends upon bases.

Having understood this, the rest is again tedious arithmetic. It is also clear that it will not be accidental if M' is related to M by the transition matrix P. Intuitively, the entire transformation process is contained in P, and every change which depends only upon the change of basis must somehow depend only upon P; it is the only "variable" in the problem. Now to the arithmetic. We follow a "programmed" procedure, inviting you to supply the missing information.

Exercise

To find M' we apply Equation (2.2) on page N38 again. This requires knowing the $\sigma(\alpha'_j)$ as linear combinations of the basis vectors

$$\boxed{\qquad\qquad} \tag{i}$$

of V. While we do not know this directly, we do know the images of the original basis vectors

$$\boxed{\qquad\qquad} \tag{ii}$$

as linear combinations of $\{\beta_1, \beta_2, \beta_3\}$, and also the new basis of U

$$\boxed{\qquad\qquad} \tag{iii}$$

as linear combinations of α_1 and α_2. Putting these together as follows yields M':

$$\sigma(\alpha'_1) = \sigma(\alpha_1 + \alpha_2)$$

$$= \boxed{\quad}\beta_1 + \boxed{\quad}\beta_2 + \boxed{\quad}\beta_3 \tag{iv}$$

and

$$\sigma(\alpha'_2) = \sigma(2\alpha_1 - \alpha_2)$$

$$= \boxed{\quad}\beta_1 + \boxed{\quad}\beta_2 + \boxed{\quad}\beta_3 \tag{v}$$

Thus

$$M' = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \tag{vi}$$

Solution

(i) $\{\beta_1, \beta_2, \beta_3\}$ (ii) $\sigma(\alpha_1), \sigma(\alpha_2)$ (iii) $\{\alpha'_1, \alpha'_2\}$

(iv) $(r_1 + s_1), (r_2 + s_2), (r_3 + s_3)$

(v) $(2r_1 - s_1), (2r_2 - s_2), (2r_3 - s_3)$

(vi) $\begin{bmatrix} r_1 + s_1 & 2r_1 - s_1 \\ r_2 + s_2 & 2r_2 - s_2 \\ r_3 + s_3 & 2r_3 - s_3 \end{bmatrix}$

Exercise

With M, P and M' as in the text above, compare MP with M'.

Solution

$$MP = \begin{bmatrix} r_1 & s_1 \\ r_2 & s_2 \\ r_3 & s_3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} r_1 + s_1 & 2r_1 - s_1 \\ r_2 + s_2 & 2r_2 - s_2 \\ r_3 + s_3 & 2r_3 - s_3 \end{bmatrix} = M'$$

The generalization of this result is given in the next reading passage.

READ page N51 *to* "... we have $A' = AP$" *(about half-way down the page).*

Note

Equation (4.6): In this equation, $Y = AX$ is the matrix representation of the action of σ, from Equation (2.9) on page N41. The next step uses Equation (4.3), and the last uses the associativity of matrix multiplication. Notice that the matrix of transition appears at the right of the product, AP.

3.1.4 General Changes of Bases for Linear Transformations

The next reading passage shows the same kind of calculation applied to changes of basis in the codomain instead of the domain. If you find the proliferating suffixes hard to interpret, especially in Equation (4.7), construct an example for yourself similar to the one at the start of the previous sub-section. For instance, use the previous example with $\{\alpha_1, \alpha_2\}$ a basis for U, σ as in Equation (1) of sub-section 3.1.3, $\{\beta_1, \beta_2, \beta_3\}$ the old basis for V and $\{\beta'_1, \beta'_2, \beta'_3\}$ a new basis for V, with

$$\beta_1 = \beta'_1 - \beta'_2$$
$$\beta_2 = 2\beta'_2 - \beta'_3$$
$$\beta_3 = \beta'_1 + \beta'_3 .$$

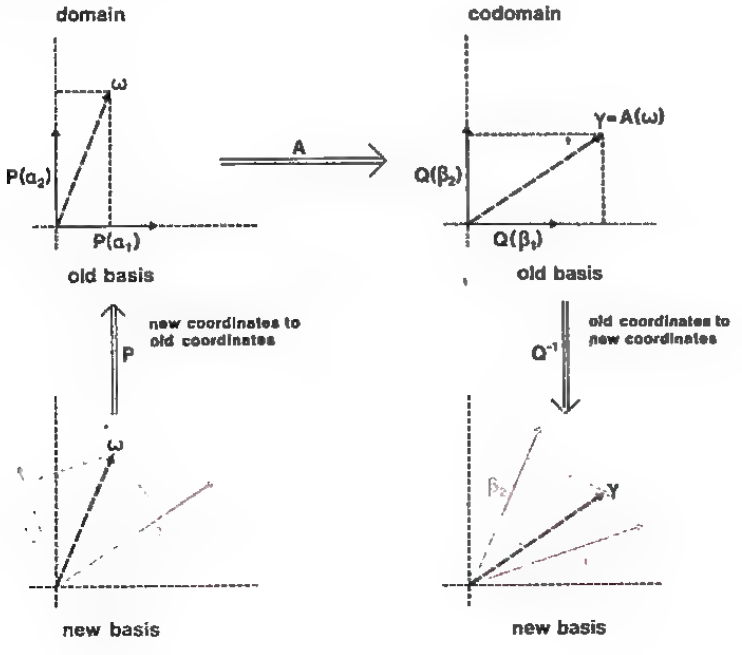
Find the transition matrix for $\{\beta'_1, \beta'_2, \beta'_3\}$ to $\{\beta_1, \beta_2, \beta_3\}$, and call it Q^{-1}. (Note that Q is the transition matrix for $\{\beta_i\}$ to $\{\beta'_i\}$.) Then find the matrix M'' representing σ with respect to $\{\alpha_1, \alpha_2\}$ and $\{\beta'_1, \beta'_2, \beta'_3\}$. Then you are on your own with Nering for the next reading passage.

READ page N51 *from* " Now consider the effect..." *to* "... σ is $Q^{-1}AP$".

Note

line −4 *page* N51 To derive the formula $Q^{-1}AP$, think of the basis changes being made in succession. We start with the matrix A with respect to the original basis. A change of basis in the domain, with fixed codomain basis, changes the matrix to AP (Equation (4.5)). Then we change the basis in the codomain, and Equation (4.8) applied to the current matrix, which is AP rather than A, tells us that the final matrix is $Q^{-1}AP$.

You may find the following interpretation helpful. In a product of linear transformations, say $\sigma\tau$, we do the right-hand one *first*, since $\sigma\tau(\alpha) = \sigma(\tau(\alpha))$. Similarly for matrices, a product $Q^{-1}AP$, which represents a linear transformation, can be interpreted as doing P first, then A, then Q^{-1}. By Equation (4.3), the matrix P takes us from the new coordinates to the old in the domain. Then A takes us from these to the codomain (using the old coordinates in both spaces). Finally, by Equation (4.3) again, to get from the old coordinates to the new in the codomain we use Q^{-1} (since Q would take us from the new to the old).



Exercises

1. For the example outlined in the first paragraph of this sub-section, find Q^{-1} and the matrix M''. Verify that $M'' = Q^{-1}M$.
2. (*Continued*) Find the matrix M''' representing σ with respect to $\{\alpha'_1 = \alpha_1 + \alpha_2, \alpha'_2 = 2\alpha_1 - \alpha_2\}$ and $\{\beta'_1, \beta'_2, \beta'_3\}$ directly. Verify that $M''' = Q^{-1}MP$.

Solutions

1. $$Q^{-1} = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

Now

$$\begin{aligned}\sigma(\alpha_1) &= r_1\beta_1 + r_2\beta_2 + r_3\beta_3 \\ &= r_1(\beta'_1 - \beta'_2) + r_2(2\beta'_2 - \beta'_3) + r_3(\beta'_1 + \beta'_3) \\ &= (r_1 + r_3)\beta'_1 + (-r_1 + 2r_2)\beta'_2 + (-r_2 + r_3)\beta'_3 \\ \sigma(\alpha_2) &= s_1\beta_1 + s_2\beta_2 + s_3\beta_3 \\ &= (s_1 + s_3)\beta'_1 + (-s_1 + 2s_2)\beta'_2 + (-s_2 + s_3)\beta'_3\end{aligned}$$

so that

$$M'' = \begin{bmatrix} r_1 + r_3 & s_1 + s_3 \\ -r_1 + 2r_2 & -s_1 + 2s_2 \\ -r_2 + r_3 & -s_2 + s_3 \end{bmatrix}$$

Also

$$Q^{-1}M = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} r_1 & s_1 \\ r_2 & s_2 \\ r_3 & s_3 \end{bmatrix}$$

$$= \begin{bmatrix} r_1 + r_3 & s_1 + s_3 \\ -r_1 + 2r_2 & -s_1 + 2s_2 \\ -r_2 + r_3 & -s_2 + s_3 \end{bmatrix}$$

2. In sub-section 3.1.3 (preceding the exercise) we found that

$$\sigma(\alpha'_1) = (r_1 + s_1)\beta_1 + (r_2 + s_2)\beta_2 + (r_3 + s_3)\beta_3$$

whence

$$\begin{aligned}\sigma(\alpha'_1) &= (r_1 + s_1)(\beta'_1 - \beta'_2) + (r_2 + s_2)(2\beta'_2 - \beta'_3) \\ &\quad + (r_3 + s_3)(\beta'_1 + \beta'_3) \\ &= (r_1 + s_1 + r_3 + s_3)\beta'_1 + (-r_1 - s_1 + 2r_2 + 2s_2)\beta'_2 \\ &\quad + (-r_2 - s_2 + r_3 + s_3)\beta'_3\end{aligned}$$

Similarly

$$\begin{aligned}\sigma(\alpha'_2) &= (2r_1 - s_1)\beta_1 + (2r_2 - s_2)\beta_2 + (2r_3 - s_3)\beta_3 \\ &= (2r_1 - s_1)(\beta'_1 - \beta'_2) + (2r_2 - s_2)(2\beta'_2 - \beta'_3) \\ &\quad + (2r_3 - s_3)(\beta'_1 + \beta'_3) \\ &= (2r_1 - s_1 + 2r_3 - s_3)\beta'_1 + (-2r_1 + s_1 + 4r_2 - 2s_2)\beta'_2 \\ &\quad + (-2r_2 + s_2 + 2r_3 - s_3)\beta'_3\end{aligned}$$

so that

$$M''' = \begin{bmatrix} r_1 + s_1 + r_3 + s_3 & 2r_1 - s_1 + 2r_3 - s_3 \\ -r_1 - s_1 + 2r_2 + 2s_2 & -2r_1 + s_1 + 4r_2 - 2s_2 \\ -r_2 - s_2 + r_3 + s_3 & -2r_2 + s_2 + 2r_3 - s_3 \end{bmatrix}$$

$$= Q^{-1}MP.$$

3.1.5 The Simplest Matrix for a Linear Transformation

By choosing suitable bases in both domain and codomain, we can bring the matrix representing the transformation to the very simple form shown on page N52. The secret, loosely speaking, is to choose a basis in the domain which has a proper subset which is a basis for the kernel of σ, and to choose a basis in the codomain which includes all the non-zero images of domain basis vectors.

Let us explain this! If the rank of σ is ρ and U is of dimension n, then $K(\sigma)$, the kernel of σ, is of dimension $n - \rho$, since by the dimension theorem, $n - \rho = \nu$, the nullity of σ. We can choose a basis of $n - \rho = \nu$ vectors for $K(\sigma)$ and then complete it to obtain a basis for U. Let the basis so obtained be

$$\overbrace{\alpha'_1, \alpha'_2, \ldots, \alpha'_\rho, \underbrace{\alpha'_{\rho+1}, \ldots, \alpha'_n}_{\text{basis for } K(\sigma)}}^{\text{basis for } U}$$

Notice that we have labelled the basis vectors of $K(\sigma)$ so that they come *last*.

As usual we write

$$\sigma(\alpha'_i) = \sum_{j=1}^{m} a_{ji}\beta_j.$$

But $\alpha'_n \in K(\sigma)$, so that $\sigma(\alpha'_n) = 0$ (the zero vector)
i.e.

$$a_{1n} = a_{2n} = \cdots = a_{mn} = 0$$

which tells us that each element of the nth column of the matrix A' representing σ is zero. Similarly, all the elements of the last ν columns of A' are zero. Notice that this is so whatever the choice of basis vectors β_i in the codomain. If $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ was the "old" basis for U with corresponding matrix A, then we can interpret our result so far, as showing that:

> If A is any $m \times n$ matrix of rank ρ, there always exists a non-singular $n \times n$ matrix P such that
>
> $$A' = AP$$
>
> has the last $n - \rho = \nu$ columns all zero.

This is the first step. The next step is to choose a "new" basis in the codomain in order to make a further simplification. We again use the dimension theorem; this time a step we noted in its proof: instead of $\beta_1, \ldots, \beta_\rho$ we use $\beta'_1, \ldots, \beta'_\rho$, where

$$\beta'_i = \sigma(\alpha_i) \qquad (i = 1, \ldots, \rho)$$

and complete to form a new basis.

READ the remainder of Section 4, pages N51–52.

Exercise

Complete the following statements to describe the steps in the proof of *Theorem 4.1*. We suggest you try to complete these steps without reference to our text or N.

(i) Given any $m \times n$ matrix A, we can regard it as representing a linear transformation σ of an n-dimensional space U to an m-dimensional space V with respect to given bases $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_m\}$ respectively.

(ii) We first choose a new basis $\{\alpha'_1, \ldots, \alpha'_n\}$ for U in the following way:

We denote the transition matrix for this change of basis by P.

(iii) We then choose a new basis $\{\beta'_1, \ldots, \beta'_m\}$ for V in the following way:

We denote the transition matrix for this change of basis by Q.

(iv) The matrix which represents σ with respect to the new bases in U and V is $Q^{-1}AP$, and its elements can be described as follows:

Solution

(ii) If $\nu = n - \rho$ is the nullity of σ, then we choose the new basis, $\{\alpha'_1, \ldots, \alpha'_n\}$, so that $\{\alpha'_{\rho+1}, \alpha'_{\rho+2}, \ldots, \alpha'_n\}$ is a basis for $K(\sigma)$.

(iii) $\beta'_i = \sigma(\alpha'_i)$ $i = 1, \ldots, \rho$; $\sigma(\alpha'_i) = 0$, $i = \rho + 1, \ldots, m$, so we choose $\beta'_{\rho+1}, \ldots, \beta'_m$ in any way that gives a basis B' of V.

(iv)

	ρ columns					ν columns
	1	0	0	$\cdots$	0	
	0	1	0	$\cdots$	0	
	.	.	.	$\cdots$	.	
ρ rows	.	.	.	$\cdots$	.	0
	.	.	.	$\cdots$	.	
	0	0	0		1	
$m - \rho$ rows			0			0

3.1.6 Summary of Section 3.1

In this section we defined the terms

matrix of transition	(page N50)
similar	(page N52)

* * *
* * *

Theorem

(*4.1, page* N52)
If A is any $m \times n$ matrix of rank ρ, there exist a non-singular $n \times n$ matrix P and a non-singular $m \times m$ matrix Q such that $A' = Q^{-1}AP$ has the first ρ elements of the main diagonal equal to 1, and all other elements equal to zero.

* *

Techniques

1. Determine the matrix of transition for a given change of basis. * * *
2. Determine the coordinates of a vector with respect to a new basis, when given the coordinates of that vector with respect to an old basis and the appropriate matrix of transition. * * *
3. Determine the matrix of a linear transformation, σ, from U to V with respect to bases A' in U and B' in V, given the matrix representing σ with respect to bases A in U and B in V, and given the form of the change of bases A to A' in U and B to B' in V. * * *

3.2 WHAT IS HERMITE NORMAL FORM?

3.2.1 Choosing a Basis in the Codomain

In the previous section we have seen that the matrix representing a linear transformation can be greatly simplified by a judicious choice of bases in the domain and codomain. Some simplification can still be achieved if we restrict ourselves to a basis change in either the domain or the codomain alone.

We stated in the Introduction, 3.0, that, from the point of view of solving systems of linear equations, basis changes in the codomain are particularly important.

For a given basis $\{\alpha_1, \ldots, \alpha_n\}$ of the domain, the choice is very simple if the transformation is an isomorphism, for then $\{\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\}$ forms a basis of the codomain and we simply take this as the new basis $\{\beta'_1, \ldots, \beta'_n\}$. (The primes over the βs indicate, as usual, that this may not have been the basis with respect to which σ was expressed originally; this would have been written $\{\beta_1, \ldots, \beta_n\}$.) Our definition for the new basis implies

$$\sigma(\alpha_1) = \beta'_1, \sigma(\alpha_2) = \beta'_2, \ldots, \sigma(\alpha_n) = \beta'_n,$$

so that the matrix of σ, with respect to these bases, is just the $n \times n$ unit matrix, I. The reason why such a simple matrix representation is possible here is that σ, being an isomorphism, has a non-singular matrix A. The change of basis we have made corresponds to using A as the matrix of transition in the codomain, so that the transformed matrix is not $Q^{-1}A$ but $A^{-1}A = I$ (by Equation (4.8), page N51).

Example 1

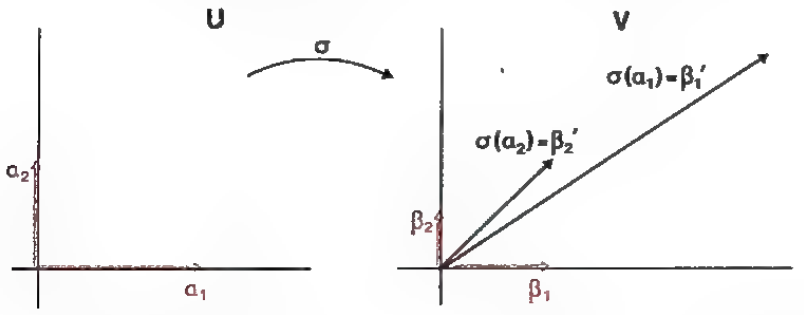
The following diagram illustrates the situation in the case where U and V are 2-dimensional, and $\sigma(\alpha_1)$ and $\sigma(\alpha_2)$ are linearly independent, and can therefore form a codomain basis $\{\beta'_1, \beta'_2\}$. We have drawn the diagram to illustrate the particular case where

$$\sigma(\alpha_1) = 3\beta_1 + 4\beta_2 = \beta'_1,$$
$$\sigma(\alpha_2) = \beta_1 + 2\beta_2 = \beta'_2.$$

The matrices representing σ with respect to the original codomain basis $\{\beta_1, \beta_2\}$ and the new codomain basis $\{\beta'_1, \beta'_2\}$ are, respectively,

$$\begin{bmatrix} 3 & 1 \\ 4 & 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

(Remember that the columns of the matrix are the coordinates of the images of the domain basis with respect to the appropriate codomain basis. Also, the original bases $\{\alpha_1, \alpha_2\}$, $\{\beta_1, \beta_2\}$ are drawn at right-angles simply to tidy up the diagram and make things easier to *visualize*. There is no mathematical significance in their being at right-angles.)



A more interesting situation arises if the images of the domain basis elements are linearly independent, but do not span the codomain.

Example 2

This example is exactly the same as the previous one, except that V is 3-dimensional, so that there is an extra codomain basis vector β_3, but we still have

$$\sigma(\alpha_1) = 3\beta_1 + 4\beta_2$$
$$\sigma(\alpha_2) = \beta_1 + 2\beta_2.$$

The matrix of σ with respect to $\{\alpha_1, \alpha_2\}$ and $\{\beta_1, \beta_2, \beta_3\}$ is now

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 2 \\ 0 & 0 \end{bmatrix}.$$

The images of the domain basis are linearly independent, just as in Example 1, but this time they do not span V. It is clear, though, that if we take

$$\beta'_1 = \sigma(\alpha_1),\ \beta'_2 = \sigma(\alpha_2),\ \beta'_3 = \beta_3,$$

then the matrix for σ is now

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

which is as simple as we can hope to make it. Furthermore, we get the same matrix whatever we take β'_3 to be (consistent with the requirement that $\{\beta'_1, \beta'_2, \beta'_3\}$ forms a basis; i.e. β'_3 is linearly independent of β'_1 and β'_2). The choice of β'_3 does not affect the fact that the images of α_1 and α_2 are expressed entirely in terms of β'_1 and β'_2. If you go back to the previous figure, and consider V to be 3-dimensional, i.e., some vectors can come out of the page at you, then you can imagine β'_3 to be any vector that does not lie in the plane of the page. You can then see that the one you choose for β'_3 does not have any effect on the relations between vectors that do lie in the plane of the page.

We shall use just the same idea in the general case. Suppose we have a transformation from an n-dimensional space U to an m-dimensional space V. In general, $\sigma(\alpha_1), \ldots, \sigma(\alpha_n)$ may fail to be linearly independent, and may also fail to span V. Suppose $\langle \sigma(\alpha_1), \ldots, \sigma(\alpha_n) \rangle$ is k-dimensional; then we shall choose our first k basis elements $\beta'_1, \ldots, \beta'_k$ in a certain way, to lie in $\langle \sigma(\alpha_1), \ldots, \sigma(\alpha_n) \rangle$, so that all of $\sigma(\alpha_1), \ldots, \sigma(\alpha_n)$ are expressible as linear combinations of these. Then we can complete to a basis $\{\beta'_1, \ldots, \beta'_m\}$ of V in any way we like, in the confidence that no matter how we choose $\beta'_{k+1}, \ldots, \beta'_m$, we have completely specified the form of the matrix by our choice of the first k basis elements. In fact, if all we want to do is to find the Hermite normal form, we do not even bother to calculate suitable $\beta'_{k+1}, \ldots, \beta'_m$; ***Theorem 3.6*** on page N17 tells us that we could do so if we felt like it, and this is enough.

Example 3

Let $U = R^2$, $V = R^2$, and $\{\alpha_1, \alpha_2\}$, $\{\beta_1, \beta_2\}$ be the standard bases.* Consider the transformation $\sigma : U \longrightarrow V$ whose matrix with respect to these bases is

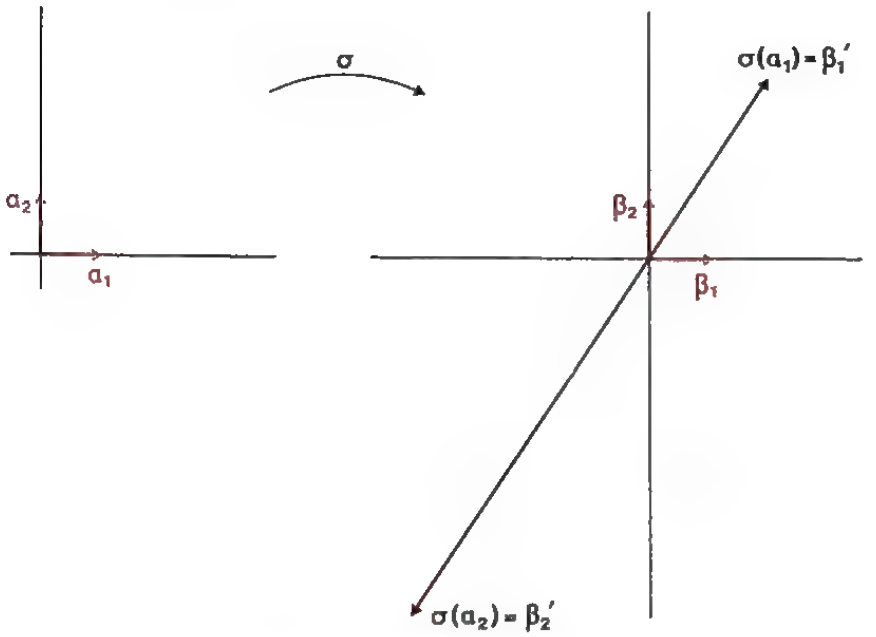
$$\begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix};$$

then $\sigma(\alpha_1) = (2, 3)$, $\sigma(\alpha_2) = (-4, -6)$, and these are linearly dependent, so they cannot both be elements of a basis for V.

* The *standard* basis for R^2 is $\{(1, 0), (0, 1)\}$.
Likewise the standard basis for R^3 is $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, and so on.

But let us take β_1' to be $\sigma(\alpha_1)$, at any rate. Do you see what has happened now? *Both* $\sigma(\alpha_1)$ *and* $\sigma(\alpha_2)$ are now expressible in terms of β_1', and once more we have completely specified the matrix. No matter how we choose β_2', we now have the matrix in the form

$$\begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix}.$$



The whole of U is mapped to the subspace of V spanned by $\beta_1' = 2\beta_1 + 3\beta_2$.

Since $\sigma(\alpha_2)$ is always dependent on $\sigma(\alpha_1)$, no matter how we choose β_1' and β_2', it would seem that we cannot make the matrix any simpler than the above by altering the codomain basis alone.

Example 4

This time let us have $U = V = R^3$, and a transformation σ whose matrix with respect to the standard bases (for both U and V) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ is

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}.$$

In this case the last column is the sum of the first two, so that

$$\sigma(\alpha_3) = \sigma(\alpha_1) + \sigma(\alpha_2).$$

That is to say, $\sigma(\alpha_1)$, $\sigma(\alpha_2)$ and $\sigma(\alpha_3)$ are linearly dependent, but *any two* of $\sigma(\alpha_1)$, $\sigma(\alpha_2)$, $\sigma(\alpha_3)$ are linearly independent. So it would seem that we should choose two of them as basis vectors, then make an arbitrary choice of the third basis vector. But which two do we choose?

It doesn't really matter exactly what rule we adopt at this point, but we must have some definite rule, and then stick to it in all cases, if we are to end up with a unique specification for Hermitè normal form. In fact, the rule which is chosen stipulates, in the case of Example 4, that $\sigma(\alpha_1)$ and $\sigma(\alpha_2)$ are to be our first two basis vectors, and then we can make an arbitrary (linearly independent) choice for the third vector.

Exercise

Write down the matrix of the transformation of Example 4, when the codomain basis is:

(i) $\beta_1' = \sigma(\alpha_1)$, $\beta_2' = \sigma(\alpha_2)$, $\beta_3' = (2, 1, 5)$;

(ii) $\beta_1' = \sigma(\alpha_1)$, $\beta_2' = \sigma(\alpha_2)$, $\beta_3' = (5, 1, 2)$.

Solution

Since $\sigma(\alpha_3) = \sigma(\alpha_1) + \sigma(\alpha_2)$, the matrix is

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

in the case of both (i) and (ii); again, the linear relationships determining the matrix are unaffected by the choice of β_3'.

Exercise

(Quite difficult, but do not spend more than a few minutes on it.)

Suggest a rule for choosing the first ρ elements of the codomain basis, when

$$\mathrm{Im}(\sigma) = \langle \sigma(\alpha_1), \ldots, \sigma(\alpha_n) \rangle$$

is ρ-dimensional, but $\sigma(\alpha_1), \ldots, \sigma(\alpha_\rho)$ are *not* linearly independent.

The solution to this exercise is contained in the text that follows.

First of all, let us look at an example where $\langle \sigma(\alpha_1), \ldots, \sigma(\alpha_n) \rangle$ is 3-dimensional, but $\sigma(\alpha_1)$, $\sigma(\alpha_2)$, $\sigma(\alpha_3)$ are not linearly independent.

Example 5

Let $U = V = R^4$, and let $\sigma : U \longrightarrow V$ have the following matrix with respect to the standard bases:

$$A = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 2 & 1 & 3 & 4 \\ 3 & 0 & 3 & 5 \\ 2 & 1 & 3 & 6 \end{bmatrix}$$

so that

$$\sigma(\alpha_1) = (1, 2, 3, 2) = \text{first column of } A,$$
$$\sigma(\alpha_2) = (2, 1, 0, 1) = \text{second column of } A,$$

and so on.

Now (2, 1, 0, 1) is not a multiple of (1, 2, 3, 2), so that $\sigma(\alpha_1)$ and $\sigma(\alpha_2)$ are linearly independent, and can presumably be taken to form the first two elements of the codomain basis for the Hermite normal form, A':

$$\beta_1' = \sigma(\alpha_1) = (1, 2, 3, 2) \tag{1}$$
$$\beta_2' = \sigma(\alpha_2) = (2, 1, 0, 1) \tag{2}$$

Now this immediately tells us what the first two columns of the new matrix are going to be. These columns, after all, are the components of $\sigma(\alpha_1)$ and $\sigma(\alpha_2)$ with respect to the new codomain basis; and *however* we choose β_3' and β_4', Equations (1) and (2) can be rewritten:

$$\sigma(\alpha_1) = 1\beta_1' + 0\beta_2' + 0\beta_3' + 0\beta_4' \tag{1a}$$
$$\sigma(\alpha_2) = 0\beta_1' + 1\beta_2' + 0\beta_3' + 0\beta_4' \tag{2a}$$

giving, for the first two columns of the new matrix A':

$$\begin{matrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{matrix}$$

What about $\sigma(\alpha_3) = (3, 3, 3, 3)$? It is not difficult to see that we have the linear relation

$$\begin{aligned} (3, 3, 3, 3) &= (1, 2, 3, 2) + (2, 1, 0, 1) \\ &= \sigma(\alpha_1) + \sigma(\alpha_2). \end{aligned} \tag{3}$$

We cannot therefore use (3, 3, 3, 3) in building up the codomain basis. In fact, Equation (3) tells us that, having fixed β'_1 and β'_2, we have automatically specified what the third column of A' must be:

$$\sigma(\alpha_3) = 1\beta'_1 + 1\beta'_2 + 0\beta'_3 + 0\beta'_4, \tag{3a}$$

so the third column must be

$$\begin{matrix} 1 \\ 1 \\ 0 \\ 0 \end{matrix}$$

Now what about $\sigma(\alpha_4) = (3, 4, 5, 6)$? If this also turns out to be dependent* on $\sigma(\alpha_1)$ and $\sigma(\alpha_2)$, then we have specified the fourth column of A', and need look no further. However, we are not so lucky this time. Note that $\sigma(\alpha_1)$ and $\sigma(\alpha_2)$ each have the second component equal to the fourth component:

$$\sigma(\alpha_1) \quad \text{is} \quad (1, \underset{\downarrow}{2}, 3, \underset{\downarrow}{2}) \quad \text{and} \quad \sigma(\alpha_2) = (2, \underset{\downarrow}{1}, 0, \underset{\downarrow}{1}).$$

So this must be true of any linear combination of $\sigma(\alpha_1)$ and $\sigma(\alpha_2)$. But it is not true of $\sigma(\alpha_4)$: for, $4 \neq 6$. Thus $\sigma(\alpha_4)$ is independent of $\sigma(\alpha_1)$ and $\sigma(\alpha_2)$, and can be taken to be a codomain basis element, β'_3.

$$\beta'_3 = \sigma(\alpha_4), \tag{4}$$

$$\sigma(\alpha_4) = 0\beta'_1 + 0\beta'_2 + 1\beta'_3 + 0\beta'_4; \tag{4a}$$

and so the fourth column of A' is

$$\begin{matrix} 0 \\ 0 \\ 1 \\ 0 \end{matrix}$$

and we have found A' without having to specify the fourth codomain basis element β'_4. This can be chosen arbitrarily; but since we have already found A', there is, in practice, no need to do so. A' is now equal to

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

It should now be apparent what the general rule must be for choosing the first ρ codomain basis elements out of $\{\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\}$ (where $\langle\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\rangle$ is ρ-dimensional). In fact, we choose them in the same way as in the proof of ***Theorem 3.5***, page N16, which you met in *Unit 1, Vector Spaces*: go through the $\sigma(\alpha_i)$ in order, picking out those which are independent of the previous ones, and rejecting those which are not. This gives us a *systematic* method of choosing the required codomain basis vectors so that they and the order in which they appear are exactly specified. We can give detailed expression to this rule as follows.

Go through each of the $\sigma(\alpha_i)$ in turn, starting with $\sigma(\alpha_1)$, then $\sigma(\alpha_2)$, etc., and accept it as a member of the basis if it is independent of all the previous ones, but reject it if it depends on the previous ones. Thus β'_1 is the first non-zero element of $\{\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\}$; β'_2 is the first element of $\{\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\}$ not depending on β'_1; β'_3 is the first one not depending on β'_1 and β'_2; and so on. The number of vectors we obtain in this way must be ρ, the rank of σ: it cannot be more, since the selection process ensures that the vectors selected are independent, and it cannot be less, since they must span the ρ-dimensional space $\langle\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\rangle$. This is because we look at each of the $\sigma(\alpha_i)$ in turn, and give it the chance of contributing something to $\langle\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\rangle$. If it adds nothing to what has gone before, i.e. if it is dependent on the previous $\sigma(\alpha_j)$'s, then and only then do we reject it.

* Here, of course, *dependent* means *linearly dependent*.

The matrix for σ obtained by choosing the first ρ elements of the codomain basis in this way, and the others arbitrarily, is the *Hermite normal form* we are looking for. Often, what interests us is the transition from the original *matrix* to the new *matrix*, so the process is usually described as *reducing a matrix to Hermite normal form.*

There is a rather neat way of characterizing what is going on here, in terms of the dimensions of the spaces spanned by sets of image vectors. This comes about as follows. We have a fixed basis $\{\alpha_1, \ldots, \alpha_n\}$ in the domain, and at the kth step in the process of choosing a basis in the codomain, we look at the first k elements $\alpha_1, \ldots, \alpha_k$ of the domain basis, and at their images $\sigma(\alpha_1), \ldots, \sigma(\alpha_k)$. Now, since $\{\alpha_1, \ldots, \alpha_k\}$ is linearly independent,

$\langle \alpha_1, \ldots, \alpha_k \rangle$ is k-dimensional.

But $\{\sigma(\alpha_1), \ldots, \sigma(\alpha_k)\}$ is not necessarily independent; what can we say about the dimension of $\langle \sigma(\alpha_1), \ldots, \sigma(\alpha_k) \rangle$?

For convenience, let

$$U_k = \langle \alpha_1, \ldots, \alpha_k \rangle.$$

Then $\langle \sigma(\alpha_1), \ldots, \sigma(\alpha_k) \rangle$ is the image of $\langle \alpha_1, \ldots, \alpha_k \rangle$, namely $\sigma(U_k)$, and the problem is to find dim $\sigma(U_k)$.

Well, it should be clear that dim $\sigma(U_k)$ must be equal to the number of elements of the codomain basis that are selected out of the set $\{\sigma(\alpha_1), \ldots, \sigma(\alpha_k)\}$ by the selection process used above. This is because, by construction, they are linearly independent, and span $\sigma(U_k)$. (The latter fact follows because those vectors $\sigma(\alpha_i)$, for $i \leqslant k$, that are *not* in this set, are linearly dependent on those which *are*.)

Thus, every time a vector $\sigma(\alpha_i)$ is chosen as a basis vector in the codomain, the dimension of $\sigma(U_i)$ becomes one more than the dimension of $\sigma(U_{i-1})$, whereas every time a vector *fails* to be chosen, the dimension of $\sigma(U_i)$ and $\sigma(U_{i-1})$ are equal. We can look at this the other way round, and *characterize* the codomain basis vectors by saying that:

$\sigma(\alpha_i)$ is a codomain basis vector if and only if

$$\dim \sigma(U_i) = \dim \sigma(U_{i-1}) + 1$$

We need another piece of notation, to help us to distinguish between those elements of $\{\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\}$ which are specially chosen to be codomain basis vectors, and those which are not. In order to do this, we give a special notation to the *suffices* corresponding to the chosen codomain basis vectors; they are denoted by $k_1, \ldots, k_\rho$. (Remember, ρ is the dimension of $\sigma(U)$, and so $\sigma(U)$ has ρ basis vectors.) Thus the corresponding *elements* of the codomain basis are denoted by $\sigma(\alpha_{k_1}), \ldots, \sigma(\alpha_{k_\rho})$ (or as $\beta'_1, \ldots, \beta'_\rho$, whichever is more convenient in the circumstances).

For instance, in Example 5 above, k_1 is 1, k_2 is 2, and k_3 is 4, so that $\sigma(\alpha_1) = \beta'_1$, $\sigma(\alpha_2) = \beta'_2$, and $\sigma(\alpha_4) = \beta'_3$.

Exercise

If the above method is applied to the matrix of Example 5, namely

$$\begin{bmatrix} 1 & 2 & 3 & 3 \\ 2 & 1 & 3 & 4 \\ 3 & 0 & 3 & 5 \\ 2 & 1 & 3 & 6 \end{bmatrix},$$

then $\sigma(U_1)$ is $\langle (1, 2, 3, 2) \rangle$, and has dimension 1.

Write down the dimensions of $\sigma(U_2)$, $\sigma(U_3)$, $\sigma(U_4)$, and specify a basis for each.

Solution

$\sigma(U_2)$ has dimension 2 and a basis $\{(1, 2, 3, 2), (2, 1, 0, 1)\}$.
$\sigma(U_3)$ has dimension 2 and the same basis as above (since $(3, 3, 3, 3) \in \sigma(U_2)$).
$\sigma(U_4)$ has dimension 3 and a basis

$$\{(1, 2, 3, 2), (2, 1, 0, 1), (3, 4, 5, 6)\}.$$

The discussion in this sub-section is an extensive elaboration of the first two paragraphs of Section 5, pages N53–54. We have not asked you to read this piece for two reasons:

(i) it refers to ***Theorem 4.8*** of Chapter I, which we have not covered;
(ii) in testing this unit, some students found the text of N particularly difficult at this point.

If you have the time you might like to look at these two paragraphs of N, to see how concisely the ideas can be put. It is a matter of taste (and possibly mathematical fashion) whether one regards such conciseness as having merit. Certainly, for those who understand it, it has great appeal, much like certain music and poetry.

3.2.2 Recognizing Hermite Normal Form

The method we used in the previous sub-section to bring a matrix to Hermite normal form can be stream-lined if we do not write out all the vectors considered, but work directly with the matrix. The method for doing this is described later in this text. To apply this method, it is necessary to be able to tell by looking at a matrix whether it is in Hermite normal form or not, and this is the subject of the present sub-section.

In Example 5 of the previous sub-section we reduced a matrix to the form

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

What features of this matrix would also be found in the Hermite normal form of a general matrix?

The columns of the above matrix are of two types. Columns 1, 2 and 4 are the images of α_1, α_2 and α_4 referred to the new codomain basis. Since these images are themselves the codomain basis vectors, these columns are

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

In the general case, the image space has dimension ρ (the rank of the matrix), there are ρ such columns, and the ith of them (which is the k_ith column in the whole matrix, if we use the notation we have just developed) consists of a solitary 1 in the ith position and zeros everywhere else.

$$\begin{matrix} \sigma(\alpha_1) & \sigma(\alpha_2) & \sigma(\alpha_3) & \sigma(\alpha_4) \\ \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ \sigma(\alpha_{k_1}) & \sigma(\alpha_{k_2}) & & \sigma(\alpha_{k_3}) \end{matrix}$$

The second type of column is exemplified by column 3 in the example. It gives the coordinates of $\sigma(\alpha_3)$, which is not included in the codomain basis because it is a linear combination of its predecessors. In general, every

column of this type is a linear combination of the columns to its left. Such a column can have any numbers in the positions with non-zero entries to the left of them, but only zeros in the positions with nothing but zeros to the left of them. Thus a general Hermite normal form matrix looks something like this:

$$\left[\begin{array}{ccc|c|ccc|c|cc|c|cc|c|cc}
0 & \cdots & 0 & 1 & \times & \times & \times & 0 & \times & \times & 0 & \times & 0 & \times & \times \\
0 & \cdots & & 0 & & & 0 & 1 & \times & \times & 0 & \times & 0 & \times & \times \\
0 & \cdots & & 0 & & & & 0 & & 0 & 1 & \times & 0 & \times & \times \\
0 & \cdots & & 0 & & & & 0 & & \cdots & 0 & & 1 & \times & \times \\
0 & & & 0 & & & & 0 & & & 0 & & 0 & 0 & 0 \\
0 & & & 0 & & & & 0 & & & 0 & & 0 & 0 & 0
\end{array}\right]$$

where the crosses can be any numbers. (The staircase-shaped line is included to guide the eye only.)

The next reading passage re-formulates the above specification in a more general way.

READ from "Since $\sigma(\alpha_{k_i}) = \beta'_i \cdots$" *on page* N54 *to the end of the statement of* ***Theorem 5.1*** *on page* N55.

Notes

(i) *First paragraph of passage.* This can perhaps be put more simply as follows. The domain basis vectors which have been labelled $\alpha_{k_1}, \ldots, \alpha_{k_\rho}$, have as their images the codomain basis vectors $\beta'_1, \ldots, \beta'_\rho$. Thus their matrix representatives, the columns $k_1, \ldots, k_\rho$, each have a 1 in the appropriate place (the ith place) and zeros elsewhere. Next, consider a typical domain basis vector α_j, with $k_i < j < k_{i+1}$. Its image is a linear combination of the codomain basis elements encountered so far, i.e. of $\beta'_1, \ldots, \beta'_i$. Thus the corresponding matrix representative has 0s below row i, and can have any numbers in the first i places (since the vector in question can be any linear combination of $\beta'_1, \ldots, \beta'_i$).
(ii) *line* -2, *page* N54 Here begins a 3-part criterion for recognizing matrices in Hermite normal form. Do not infer from the wording of part (1) of the criterion that it is necessary to know the rank ρ before we can apply the criterion. On the contrary, if all 3 parts of the criterion are satisfied for some value of ρ, then that value of ρ must be the rank.
(iii) ***The last line of the statement of Theorem 5.1.*** This is dealt with in sub-section 3.2.3.

Exercise

Do Exercise 1 on page N56, and determine the ranks of those matrices that are in Hermite normal form.

Solution

Matrices (a), (c) and (d) are in Hermite normal form, and they all have rank 3.

Matrix (b) violates part (2) of the criterion on page N55, because the first non-zero element in row 1 is not a 1.

Matrix (e) violates part (3) of the criterion, because (with $k_1 = 1$, $k_2 = 2$, $k_3 = 4$, $k_4 = 5$) there is a non-zero element in column 5 other than the 1 in row 4.

3.2.3 The Uniqueness of Hermite Normal Form

Reviewing what we have done so far in Section 3.2:

(i) We have seen that choosing a suitable codomain basis can greatly simplify the matrix of a transformation.
(ii) We have seen how to lay down a rule for choosing the codomain basis, which uniquely specifies a form for a particular matrix.
(iii) We have specified the conditions which a matrix obeys if it is in this form, namely Hermite normal form.

What we have *not* done, is to prove the *uniqueness* of Hermite normal form. There might, for all we know so far, be a different codomain basis choice, giving a different form for the matrix, but still obeying the three conditions laid down in ***Theorem 5.1***, pages N54–55, for determining Hermite normal form. The proof that this is *not* so, and that there is a *unique* matrix in Hermite normal form to be obtained from any given matrix by a change of codomain basis, is given in the next reading passage. You can consider this proof as a test for yourself; if you can follow it quite easily, then you have probably grasped the material so far. If you cannot follow the proof, then read again Section 3.2 up to this point, and go over the exercises again.

READ page N55 *from* "*The form A′ is . . .*" *as far as but excluding* **Definition.**

Notes

(i) *line 19, page* N55 Condition (3) tells us that the k_ith column, which gives the image of α_{k_i} in the new basis, has a 1 as its ith entry and zeros for all the others, so that this image is the ith codomain basis vector both in B_1' and B_2'.
(ii) *line 20, page* N55 Condition (1) tells us that every column has only zeros for the $(\rho + 1)$th entry onwards, so that all the images of the domain basis vectors are linear combinations of $\{\sigma(\alpha_{k_1}), \ldots, \sigma(\alpha_{k_\rho})\}$. The coefficients in these linear combinations are the non-zero entries in the columns of the matrix, which are therefore the same both for B_1' and B_2'.

Exercise

If A, B and M are matrices, with M non-singular, such that $A = MB$, what can be deduced about the Hermite normal forms of A and B?

Solution

The Hermite normal forms of A and B are identical.

(*Proof **Theorem 5.1*** tells us that there is a non-singular matrix Q such that $Q^{-1}A = A'$ is the unique Hermite normal form (HNF) of A. Since $A = MB$, we have

$$(Q^{-1}M)B = A',$$

and since $Q^{-1}M$ is non-singular and A' is in HNF, it follows, by the uniqueness of HNF, that A' is also the HNF of B.)

3.2.4 Row and Column Rank; the Transpose

This sub-section is just to draw your attention to ***Proposition 5.2***, at the bottom of page N55.

The proof of ***Proposition 5.2*** involves the concept of the *transpose* of a matrix. It is not a difficult concept to understand or remember. For example, the transpose of

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \text{is} \quad \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

READ the definition at the bottom of page N55.

The next exercises develop certain important properties of the transpose. It is worth spending ten or fifteen minutes on them if necessary, before you look up the answers. If you find the exercises difficult, try some special cases (with numbers for the matrix elements) first.

Exercises

1. Show that $(A^T)^T = A$; i.e. that the transpose of the transpose of A is A.
2. Exercise 4, page N57.

Solutions

1. To find a given entry in the transpose, one simply swaps the two suffices around. Doing this twice simply gets back to the original entry. Thus $(A^T)^T = A$.

   This is a case where the argument is almost trivial, but it is a little more difficult to find suitable notation with which to write out the argument in symbols. If we stick to the convention of using lower-case letters for matrix entries, then we can denote the (ij)th entry in A^T by a'_{ij}, and the (ij)th entry in $(A^T)^T$ by a''_{ij}. Then:

   $$\begin{aligned} a''_{ij} &= a'_{ji} && \text{(for all relevant } i, j) \\ &= a_{ij} && \text{(for all relevant } i, j) \end{aligned}$$

   Thus $(A^T)^T = A$.

2. (a) Let $A + B = C$, and denote the (ij)th entries of A, A^T, B, B^T, C, C^T by a_{ij}, a'_{ij}, etc. Then

   $$\begin{aligned} c'_{ij} &= c_{ji} && \text{(for all relevant } i, j) \\ &= a_{ji} + b_{ji} && \text{(for all relevant } i, j) \\ &= a'_{ij} + b'_{ij} && \text{(for all relevant } i, j) \end{aligned}$$

   Thus $C^T = A^T + B^T$.

   (b) Let $AB = C$, and use the same notation convention as above. For AB to be defined, A must be $(m \times n)$, and B must be $(n \times p)$, for some $m, n, p \in Z^+$. Then C is $(m \times p)$. It follows that A^T, B^T and C^T are $(n \times m)$, $(p \times n)$, $(p \times m)$ respectively. We then have

   $$\begin{aligned} c'_{ij} &= c_{ji} && (i = 1, \ldots p, j = 1, \ldots m) \\ &= \sum_{k=1}^{n} a_{jk} b_{ki} \\ &= \sum_{k=1}^{n} a'_{kj} b'_{ik} \\ &= \sum_{k=1}^{n} b'_{ik} a'_{kj} && \text{(just putting the previous line in a more recognizable form)} \end{aligned}$$

   This shows that $C^T = B^T A^T$.

(c) $AA^{-1} = I$

Therefore

$$(A^{-1})^T A^T = (AA^{-1})^T \qquad \text{(by (b) above)}$$
$$= I^T = I.$$

Thus $(A^{-1})^T$ is the inverse of A^T, i.e.

$$(A^{-1})^T = (A^T)^{-1}$$

"The transpose of the inverse is the inverse of the transpose."

READ ***Proposition 5.2*** *and its proof on pages* N55–56.

Notes

(i) *line 2, page* N56 Note the use of ***Theorem 3.7*** from page N48, namely that multiplication by a non-singular matrix (here Q^{-1}) does not change the rank of a matrix. The result is used again in line 5.
(ii) *line 3, page* N56 While we generally distrust statements such as "it is obvious" (two weeks' work), "the student will easily verify that" (a Ph.D dissertation), "it is left as an exercise for the serious student" (a problem my Ph.D supervisor never got around to solving for me), in this particular case it should be pretty obvious that the rank of $(A')^T$ is ρ. A perusal of the matrix labelled (5.1) on page N54 shows that all entries below the ρth row must be zero, so the rank of $(A')^T$ is less than or equal to ρ. Further, if $j \leqslant \rho$, the jth row contains, as its k_jth entry, the only non-zero entry in the k_jth column, so that it cannot be a linear combination of the other rows. Thus, all of the first ρ rows of A' are linearly independent; i.e., the rank of $(A')^T$ is ρ.

Example

Consider the following matrix in Hermite normal form:

$$A' = \begin{array}{c} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} \begin{array}{c} \begin{array}{cccccccc} k_1 & k_2 & & & & k_3 & & \end{array} \\ \begin{bmatrix} \textcircled{1} & 0 & -2 & 2 & 0 & 0 & 1 \\ 0 & \textcircled{1} & 1 & -4 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & \textcircled{1} & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

There are three columns corresponding to codomain basis elements: columns k_1, k_2, k_3 shown. Thus the number of linearly independent columns is 3. As for the rows, only the first three are non-zero, so there are a maximum of 3 linearly independent rows. But the circled 1s in rows 1, 2 and 3 come in columns that are otherwise composed entirely of zeros. Thus none of these rows can possibly be expressed as a linear combination of the other rows. So 3 is the *exact* number of linearly independent rows of A', and is therefore the *exact* number of linearly independent columns of $(A')^T$, which is therefore of rank 3. For convenience, we write out $(A')^T$ below:

$$(A')^T = \begin{array}{c} \\ k_1 \\ k_2 \\ \\ k_3 \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccc} & & & & & \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \textcircled{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \textcircled{1} & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & 0 \\ 2 & -4 & 0 & 0 & 0 & 0 \\ 0 & 0 & \textcircled{1} & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 1 & 2 & 3 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

3.2.5 Summary of Section 3.2

In this section we defined the terms

standard basis for R^n	(page C20)*	* * *
Hermite normal form	(page N55)	* * *
transpose of a matrix	(page N55)	* * *

We introduced the notation

U_k	(page N54)
$k_1, k_2, \ldots, k_\rho$	(page N54)
A^T	(page N55)

Theorems

1. (*5.1*, page N54) * * *
   Given any $m \times n$ matrix A of rank ρ, there exists a non-singular $m \times m$ matrix Q such that $A' = Q^{-1}A$ has the following form:
   (1) There is at least one non-zero element in each of the first ρ rows of A' and the elements in all remaining rows are zero.
   (2) The first non-zero element appearing in row $i (i \leqslant \rho)$ is a 1 appearing in column k_i, where $k_1 < k_2 < \cdots < k_\rho$.
   (3) In column k_i the only non-zero element is the 1 in row i.
2. (***Proposition 5.2***, page N55) * * *
   The number of linearly independent rows in a matrix is equal to the number of linearly independent columns.

Techniques

1. Recognize when a matrix is in Hermite normal form, and hence determine the rank of such a matrix. * * *
2. Given the matrix, A, representing a linear transformation $\sigma : U \longrightarrow V$ with respect to specified bases in U and V, determine the Hermite normal form of A by selecting a linearly independent subset of $\{\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\}$, $\{\alpha_1, \ldots, \alpha_n\}$ being the specified basis in U. * *
3. Use the properties of the transpose of a matrix in matrix manipulations. * * *

* C20 means page 20 of this correspondence text.

3.3 HOW TO CALCULATE HERMITE NORMAL FORM

3.3.1 Elementary Operations and Elementary Matrices

This section covers Chapter II, Section 6 of N, which gives a method by which Hermite normal form can be calculated in practice. In order to prove the viability of the concept of Hermite normal form, we constructed the Hermite normal form for a matrix by a method which, though useful for demonstrating the existence and uniqueness of Hermite normal form, gave no practical method for finding the Hermite normal form for an arbitrary matrix. Our method, remember, was to consider the matrix to be the matrix of a linear transformation $\sigma : U \longrightarrow V$, with respect to a basis $\{\alpha_1, \ldots, \alpha_n\}$ of U and $\{\beta_1, \ldots, \beta_m\}$ of V. We looked through the images of the domain basis, and applied the rules:

(i) If $\sigma(\alpha_i)$ is linearly independent of $\sigma(\alpha_1), \ldots, \sigma(\alpha_{i-1})$, take $\sigma(\alpha_i)$ as a basis element of the codomain.

(ii) If $\sigma(\alpha_i)$ is linearly dependent on $\sigma(\alpha_1), \ldots, \sigma(\alpha_{i-1})$, then express it in terms of the basis elements $\sigma(\alpha_{k_1}), \ldots, \sigma(\alpha_{k_s})$ which have already been chosen.

However, in the examples we looked at, there was always a rather easy "ad hoc" method available of seeing whether $\sigma(\alpha_i)$ was a linear combination of $\sigma(\alpha_1), \ldots, \sigma(\alpha_{i-1})$, and if so, exactly what the linear combination was. This would not be true in general, and we have not yet given a general method of answering these questions. It would therefore seem at first sight that we have to find such a method, then apply it in order to find a Hermite normal form. Actually, as we shall see presently, it is easier to go the other way round: Section 6 of Chapter II in N gives a method of obtaining Hermite normal form, which is then immediately applicable to finding the linear relations which exist among a set of vectors.

This method relies on *elementary row operations*, which you met in *Unit M100 26, Linear Algebra III*. Basically, these operations permute the rows of a matrix, or change a given row into a linear combination of itself and other rows. Such operations are (as you will shortly see) equivalent to multiplying the matrix on the left by non-singular matrices, which are (see Equation (4.8) on page N51) in turn equivalent to changing the codomain basis with respect to which the matrix is expressed (sub-section 3.1.4). Successive applications of these operations get the matrix closer and closer to Hermite normal form, and the fact that each step is a left-multiplication by a non-singular matrix ensures that the codomain basis is changed into a new codomain basis each time, with the domain basis kept fixed. So we end up with a matrix which is indeed the matrix of the original linear transformation, but expressed with respect to a new codomain basis.

READ Section 6, page N57 *up to and including the statement of* ***Theorem 6.1****, page* N58.

Exercises

1. Try to prove the statement in the sentence immediately preceding ***Theorem 6.1.***
2. Invent an example of a 3×3 matrix A. Apply three different elementary row operations in succession to it. Then apply the same operations to the unit 3×3 matrix, and check that multiplication of A on the left by this matrix has the same result as the application of the operations to A.
3. If A is an $m \times n$ matrix, to which unit matrix would you apply an elementary row operation, in order to obtain the requisite elementary matrix? (Remember, a unit matrix can be 1×1, 2×2, 3×3, etc.)

Solutions

1. If E is the elementary matrix representing a given elementary row operation, then the effect of that operation on any matrix A is, by definition, to transform it into the matrix EA. In particular, this is true of the unit matrix I, which is transformed into the matrix

$$EI = E.$$

Thus E is the matrix obtained by performing the elementary row operation on I.
In exactly the same way, the matrix representing a whole sequence of elementary row operations can be obtained by performing this sequence of operations, in the right order, on the unit matrix.

2. The following is one example:

$$\text{Let } A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 2 & 4 & 6 \end{bmatrix}.$$

Consider the effect of applying the following elementary row operations, in the following order:

(i) Interchange rows 1 and 2
(ii) Multiply row 3 by $\frac{1}{2}$
(iii) Add (-1) times row 2 to row 3.

After operation (i), the matrix is $\begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$.

After operation (ii), the matrix is $\begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$.

After the final operation, the matrix is $\begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$.

If we apply the same operations, in the same order, to the unit matrix, we obtain

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & \frac{1}{2} \end{bmatrix}.$$

It is easy to check that

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 2 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}.$$

3. It is $m \times m$, because left-multiplication of A can only be defined for a $p \times m$ matrix.

3.3.2 The Use of the Elementary Operations

READ the rest of Section 6 starting on page N58. (You should find the worked examples useful.)

Note

Page N61. As N says in the final paragraph on this page, the method described is the easiest available hand computation method for finding the inverse of a non-singular matrix.

The following exercises will give you practice in using the algorithm described in the reading passage.

If you wish to use the computer program $MPM2ϕ1, parts (c) and (d) of Exercise 1, (c) of Exercise 2, and Exercise 3, are probably worth doing this way. Save up all the exercises in this text which you want to do by $MPM2ϕ1 for one trip to the computer terminal.

Exercises

1. Obtain the Hermite normal form of the following matrices.

   (a) $\begin{bmatrix} 2 & 1 & 1 \\ 2 & 2 & 3 \\ 2 & 1 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} -2 & 1 & -3 \\ 1 & 1 & 3 \\ 0 & 1 & 1 \end{bmatrix}$

   (c) $\begin{bmatrix} 2 & 1 & 0 & 0 & 3 \\ 2 & 2 & 0 & 0 & 5 \\ 0 & 0 & 0 & 1 & 1 \\ 4 & 3 & 0 & 0 & 8 \end{bmatrix}$

   (d) $\begin{bmatrix} 2 & 4 & 6 & 6 \\ 1 & 2 & 3 & 5 \\ 3 & 6 & 9 & 0 \\ 2 & 4 & 6 & 2 \end{bmatrix}$

2. Find the inverses of the following matrices, using the method described in the last paragraph on page N61.

   (a) $\begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \end{bmatrix}$

3. (Optional)
   This exercise is designed to show you something about the limitations of using a computer for matrix operations. The object is to invert a matrix whose rows are nearly, but not quite, linear combinations of each other*. A relatively small "round-off" error in the initial entries leads to a large error in the result.

   The matrix in question is $\begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \end{bmatrix}$. It is often called the Hilbert matrix.

   Its true inverse is $\begin{bmatrix} 72 & -240 & 180 \\ -240 & 900 & -720 \\ 180 & -720 & 600 \end{bmatrix}$.

   Evaluate the inverses of

   (a) $\begin{bmatrix} 0.5 & 0.333333 & 0.25 \\ 0.333333 & 0.25 & 0.2 \\ 0.25 & 0.2 & 0.166667 \end{bmatrix}$

   and (b) $\begin{bmatrix} 0.5 & 0.333 & 0.25 \\ 0.333 & 0.25 & 0.2 \\ 0.25 & 0.2 & 0.167 \end{bmatrix}$, and compare them.

* Such a problem is said to be ill-conditioned (see *Unit 8*).

Solutions

1. (a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$

   (b) $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

   (c) $\begin{bmatrix} 1 & 0 & 0 & 0 & 0.5 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

   (d) $\begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

2. (a) $\begin{bmatrix} -1 & 1 \\ 3 & -2 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} -2 & 0 & 1 \\ 0 & 3 & -2 \\ 1 & -2 & 1 \end{bmatrix}$

3. (a) You should get something like

$$\begin{bmatrix} 71{\cdot}9787 & -239{\cdot}92 & 179{\cdot}935 \\ -239{\cdot}92 & 899{\cdot}699 & -719{\cdot}757 \\ 179{\cdot}935 & -719{\cdot}755 & 599{\cdot}802 \end{bmatrix}$$

   quite close to the true inverse.

   (b) You should get something like

$$\begin{bmatrix} 55{\cdot}49 & -177{\cdot}917 & 130{\cdot}005 \\ -177{\cdot}917 & 665{\cdot}881 & -531{\cdot}12 \\ 130{\cdot}005 & -531{\cdot}118 & 447{\cdot}439 \end{bmatrix}$$

   which is nothing like the true inverse.

   Inverting the Hilbert matrix and other similar problems are discussed in *Unit 8, Numerical Solution of Simultaneous Algebraic Equations*

3.3.3 Summary of Section 3.3

In this section we defined the terms

elementary row operation	(page N57)	* * *
elementary matrix	(page N58)	* * *

We introduced the notation

$E_2(c)$, $E_{31}(k)$, E_{12}, etc.	(page N57–58)

Theorem

(*6.1*, page N58)
Any non-singular matrix A can be written as a product of elementary matrices. * * *

Techniques

1. Use row operations to reduce a matrix to Hermite normal form. * * *
2. Use row operations to find the inverse of a matrix. * * *

3.4 APPLICATIONS OF HERMITE NORMAL FORM

3.4.1 Linear Problems and Linear Equations

Systems of linear equations, such as

$$\begin{aligned} 4x_1 + 3x_2 + 2x_3 - x_4 &= 4 \\ 5x_1 + 4x_2 + 3x_3 - x_4 &= 4 \\ -2x_1 - 2x_2 - x_3 + 2x_4 &= -3 \\ 11x_1 + 6x_2 + 4x_3 + x_4 &= 11 \end{aligned} \qquad (1)$$

arise in many applications of mathematics, of which structural engineering, management theory, and electric circuit theory are a few examples. You have already seen, in *Unit M100 26, Linear Algebra III*, how to solve such equations by the Gauss elimination method in the case where the matrix formed by the coefficients on the left is non-singular; but in this particular case the matrix *is* singular, and the Gauss elimination method breaks down. In this section we see how the Hermite normal form can be used to solve any linear problem, whether the linear transformation involved is singular or non-singular. (It is used to solve the system (1) on pages N65–66.) The only restriction is that the vector spaces involved must be of finite dimension, so that the linear transformation has a matrix representation.

The next passage from N reviews the general features of linear problems, which you have seen in the Foundation Course, and proceeds to prove a theorem which is useful for deciding whether or not a linear problem has a non-empty solution set. It then proceeds to show the relevance of Hermite normal form to linear problems. Basically, if one reduces the matrix associated with the problem to Hermite normal form, then one ends up with a new problem which is generally much simpler to solve, and has exactly the same solution set as the original problem.

The theorem proved in this passage is a result which you met in *Unit M100 26*; you should know it and be able to prove it.

READ Section 7, from page N63, *as far as but excluding* ***Theorem 7.2****, page* N66.

Notes

(i) *Equation (7.1), page* N63 $\{\xi_0\} + K(\sigma)$ means the set of all vectors of the form $\gamma + \delta$ with $\gamma \in \{\xi_0\}$ (i.e. $\gamma = \xi_0$) and $\delta \in K(\sigma)$.
(ii) *lines 4, 5, page* N64 If $\beta \neq 0$, then the solution set cannot be a subspace, because it cannot contain the zero vector: $\sigma(0) = 0 \neq \beta$. It is a slight stretching of the word "dimension" to talk of the dimension of the solution set in this case, but it is obvious what is meant.
(iii) *line 6, page* N64 Remember that the nullity is the rank of $K(\sigma)$.
(iv) *line 12, page* N64 The notation $(b_1, \ldots, b_m)$ for $m \times 1$ column matrices was introduced on page N42.
(v) ***Theorem 7.1****, page,* N64 The statement "all solutions can be expressed in terms of ν independent parameters" means that the solution set will be of the form

$$\{\xi = \xi_0 + x_1\alpha_1 + \cdots + x_\nu\alpha_\nu\}$$

where the "parameters" $x_1, \ldots, x_\nu$ take all values in $\mathbf{F}$; i.e. every ν-tuple $(x_1, \ldots, x_\nu)$ gives an element of the solution set. (In the above expression, ξ_0 is a particular solution and $\{\alpha_1, \ldots, \alpha_\nu\}$ is a basis of $K(\sigma)$.)
(vi) *line 1, page* N65 In reading this paragraph, particularly Equation (7.5), it will help if you refer back to the general Hermite normal form (5.1) on page N54.

One further point which needs a little amplifying is the definition of an augmented matrix; it is easy enough to see how the augmented matrix is constructed, but just what sort of a linear transformation does it represent? The answer is that, though one can of course define a linear transformation whose matrix is the augmented matrix, there is no particular point in

doing so—the transformation so obtained has no particular significance for the problem. This is an illustration of the fact that a matrix is not exactly the same thing as a linear transformation. It can be used to represent a linear transformation, but it can be put to other uses as well. Other cases where a matrix is used which does not directly represent a linear transformation, are:

(a) The "augmenting" of a matrix A to $[A, I]$

$$= \begin{bmatrix} a_{11} \cdots a_{1n} & 1 & 0 \cdots 0 \\ a_{21} \cdots a_{2n} & 0 & 1 \cdots 0 \\ \vdots \quad\quad \vdots & \vdots & \vdots \quad \vdots \\ a_{n1} \cdots a_{nn} & 0 & 0 \cdots 1 \end{bmatrix}$$

when using the Hermite normal form method to find the inverse of a matrix (see page N61).

(b) The use of the *row-sum check* when finding Hermite normal form (see also *Unit M100 26*). This is a way of checking the correctness of the arithmetic in row operations; the idea is to include an extra column, each entry of which is the sum of all the entries in the corresponding row. For example, if you want to find the Hermite normal form of

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 0 \\ 1 & 1 & 2 & 2 \end{bmatrix}$$

you should include an extra column consisting of the row sums of the original matrix, to get

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 10 \\ 3 & 2 & 1 & 0 & 6 \\ 1 & 1 & 2 & 2 & 6 \end{bmatrix}$$

since $1 + 2 + 3 + 4 = 10$, etc.

You then find the Hermite normal form for this matrix. Since row operations do not disturb existing linear relations among columns, when you have finished, the last column should still consist of row sums. In the above case, if you perform the operations and get

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 2 \end{bmatrix}$$

then you can verify that $1 + 0 + 0 - 1 = 0$, $0 + 1 + 0 + 1 = 2$, etc., so that it is *probable* that no arithmetical blunders have been made.

Dropping the check column gives the Hermite normal form of the original matrix to be

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

On the other hand, if at any stage of the calculation the check column is *not* equal to the sum of the others, then you know there is a blunder somewhere, and you can rectify it before it does any damage.

For example, suppose you start with

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 10 \\ 3 & 2 & 1 & 0 & 6 \\ 1 & 1 & 2 & 2 & 6 \end{bmatrix}$$

as before, and carry out the operations

row 2 − 3 row 1 and row 3 − row 1

to get

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 10 \\ 0 & -4 & -8 & -12 & -24 \\ 0 & 1 & -1 & -2 & -4 \end{bmatrix}$$

Then since $1 - 1 - 2 \neq -4$, there must be a mistake somewhere in the last row.

Exercises

1. Exercise 2, page N67, using the row sum check in the computations.
2. Exercise 4, page N68 } using either the computer or
3. Exercise 5, page N68 } the row-sum check (or both!)
4. Exercise 7, page N68.

Solutions

1. The augmented matrix, with row sums as well, is

$$\begin{array}{c} \\ \begin{bmatrix} 1 & 2 & -3 & 1 & 0 & 1 \\ 3 & -1 & 5 & -1 & 0 & 6 \\ 2 & 1 & 0 & 1 & 0 & 4 \end{bmatrix} \end{array} \quad \begin{array}{c} \text{row} \\ \text{sums} \end{array}$$

Its Hermite normal form is

$$\begin{array}{c} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 2 \\ 0 & 1 & -2 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \end{array} \quad \begin{array}{c} \text{row} \\ \text{sums} \end{array}$$

corresponding to the equations

$$\begin{aligned} x_1 + x_3 &= 0 \\ x_2 - 2x_3 &= 0 \\ x_4 &= 0. \end{aligned}$$

The only unknown to appear in more than one equation is x_3. This can be given an arbitrary value, and the above equations then fix x_1, x_2 and x_4 in terms of x_3.

The solution can be written in the form

$$\begin{aligned} x_1 &= -x_3 \\ x_2 &= 2x_3 \\ x_3 &= x_3 \\ x_4 &= 0 \end{aligned}$$

with x_3 arbitrary (compare the top of page N66).

The space of solutions is $\langle(-1, 2, 1, 0)\rangle$. Incidentally, this space is the kernel of the corresponding linear transformation.

2. The augmented matrix is

$$\begin{bmatrix} 2 & -1 & -3 & 1 \\ 1 & -1 & 2 & -2 \\ 4 & -3 & 1 & -3 \\ 1 & 0 & -5 & 3 \end{bmatrix}$$

whose Hermite normal form is

$$\begin{bmatrix} 1 & 0 & -5 & 3 \\ 0 & 1 & -7 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

corresponding to the equations

$$x_1 - 5x_3 = 3$$
$$x_2 - 7x_3 = 5.$$

Again x_3 is the only unknown appearing in more than one equation, and so we obtain

$$x_1 = 3 + 5x_3$$
$$x_2 = 5 + 7x_3$$
$$x_3 = \quad x_3$$

with x_3 arbitrary. This can also be written

$$(x_1, x_2, x_3) = (3, 5, 0) + x_3(5, 7, 1)$$

and so the solution set (not a subspace this time) is

$$(3, 5, 0) + \langle(5, 7, 1)\rangle.$$

3. The augmented matrix is

$$\begin{bmatrix} 7 & 3 & 21 & -13 & 1 & -14 \\ 10 & 3 & 30 & -16 & 1 & -23 \\ 7 & 2 & 21 & -11 & 1 & -16 \\ 9 & 3 & 27 & -15 & 1 & -20 \end{bmatrix}$$

whose Hermite normal form is

$$\begin{bmatrix} 1 & 0 & 3 & -1 & 0 & -3 \\ 0 & 1 & 0 & -2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

corresponding to the equations

$$x_1 + 3x_3 - x_4 = -3$$
$$x_2 - 2x_4 = 2$$
$$x_5 = 1.$$

In this case, x_3 and x_4 can be chosen arbitrarily; we rewrite the equations

$$x_1 = -3 - 3x_3 + x_4$$
$$x_2 = 2 + 2x_4$$
$$x_5 = 1$$

or

$$(x_1, x_2, x_3, x_4, x_5) = (-3, 2, 0, 0, 1) + (-3, 0, 1, 0, 0)x_3 + (1, 2, 0, 1, 0)x_4$$

giving the solution set

$$(-3, 2, 0, 0, 1) + \langle(-3, 0, 1, 0, 0), (1, 2, 0, 1, 0)\rangle.$$

4. The associated homogeneous problem is

$$\frac{d^2y}{dx^2} + 4y = 0.$$

Its solution is

$$y = C_1 \sin 2x + C_2 \cos 2x.$$

The most obvious particular solution is

$$y = \tfrac{1}{3} \sin x.$$

3.4.2 Linear Relations Among a Given Set of Vectors

Example 1

A brewer uses hops, barley, sugar and water to produce various types of beer. He uses them in the following proportions (by weight):

Beer	Hops	Barley	Sugar	Water
Mild	1	10	20	200
Ordinary Bitter	3	15	20	200
'Soopa-Broo'	5	20	20	200

The prices of the ingredients vary throughout the year, and in calculating the cost to him of producing a barrel of each type of beer, he laboriously goes through a calculation for each of the three types, once a week. This is all quite complicated, as the price of hops is given per ounce, of barley per bushel, of sugar per pound and of water per thousand gallons. (He looks forward enthusiastically to metrication.)

Can we help him to save some labour? The answer is yes: the vectors (1, 10, 20, 200), (3, 15, 20, 200) and (5, 20, 20, 200), representing the proportions of ingredients in each type of beer, are linearly dependent. Calling them $\beta_1, \beta_2, \beta_3$ respectively, we in fact have

$$\beta_2 = \tfrac{1}{2}(\beta_1 + \beta_3).$$

Since the cost of producing the beer depends linearly on the price of the ingredients, it follows that he only has to go through two of his complicated calculations, for Mild and "Soopa-Broo" say. Ordinary Bitter will then always be the average of the prices of Mild and "Soopa-Broo".

The general case is not quite so easy. There may be several vectors, and one may not know which vectors to try to express as linear combinations of which others. Even in the case of three vectors, the situation may be more complicated than the one considered above.

Example 2

Consider the three vectors in R^4

$$\beta_1 = (3, 0, 5, 8), \beta_2 = (1, 2, 3, 4), \beta_3 = (1, -1, 1, 2).$$

It is not at all obvious whether β_3 is a linear combination of β_1 and β_2 and if so, just what linear combination it is. The secret of the method is to write the components of the given vectors as columns of a matrix

$$B = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 2 & -1 \\ 5 & 3 & 1 \\ 8 & 4 & 2 \end{bmatrix}$$

and to use the fact that elementary row operations on B will not alter whatever linear relations exist among these columns*. This fact tells us in particular that the row operations that convert B to Hermite normal form will not alter these linear relations.

* *Proof* Such a linear relation could be written in the form

$$\begin{bmatrix} 3 & 1 & 1 \\ 0 & 2 & -1 \\ 5 & 3 & 1 \\ 8 & 4 & 2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \text{ i.e. } BC = 0.$$

Elementary row operations convert B to $B' = EB$ where E is non-singular, and hence we have $B'C = EBC = 0$ if and only if $BC = 0$.

The Hermite normal form of B is

$$\begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and nothing could be easier than to read off from this that the first two columns of B are linearly independent, and that the third is half the first minus half the second; in other words, that

$$\beta_3 = \tfrac{1}{2}\beta_1 - \tfrac{1}{2}\beta_2 .$$

The same method can be applied to any set of vectors in any finite dimensional space, provided that they are specified by their coordinates with respect to some basis. This general case is discussed further in the next reading passage. The first part of the passage (down to "an exercise for the reader" on page N73) is more complicated than the above treatment, because it deals explicitly with the change of basis implied by converting B to Hermite normal form, instead of considering only the matrix as we have done. Do not spend a lot of time working through this part of the argument in N; the most important thing here is not the theory, but how to do the calculation.

READ from "Let $B = \ldots$", *page* N72 *to the end of Section 8, page* N74.

Notes

(i) *Equations (8.2), (8.3), page* N72 These give the vectors $\beta_1, \ldots, \beta_n$ as linear combinations of $\{\alpha_1, \ldots, \alpha_m\}$ (Equation (8.2)) and $\{\alpha'_1, \ldots, \alpha'_m\}$ (Equation (8.3)). By labelling the coordinates with suffices in the manner described, matrices A and A' are constructed from the coordinates in such a way that each column gives the coordinates of the corresponding element of the βs.
(ii) *line 4, page* N73 We have so far been talking in general terms, and have not actually specified $\alpha'_1, \ldots, \alpha'_m$. We now specify that these vectors shall be chosen in such a way that the matrix A' is the Hermite normal form of A. That is to say, we select a linearly independent set from the vectors $\beta_1, \ldots, \beta_n$. As in the reading passage on page N54, we denote the suffices of these distinguished vectors by $k_1, k_2, \ldots$.

Exercises

1. For each of the following sets of vectors, find a linearly independent subset, and express any vectors which are not in this subset in terms of those which are.

   (a) (1, 2, 3), (3, 2, 1), (1, 1, 2)
   (b) (1, 2, 3), (3, 2, 1), (1, 1, 1)
   (c) (1, −1, 2), (2, −2, 4), (0, 1, 2), (1, $\frac{1}{2}$, 5)

   Remember to keep a row-sum check in your calculation, if you do it by hand.
2. Express (1, 3, 6, 8) as a linear combination of (2, 3, 4, 5), (1, 2, 3, 4), (0, 1, 1, 0) and (0, 0, 1, 1).

Solutions

1. (a) The Hermite normal form of $\begin{bmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \\ 3 & 1 & 2 \end{bmatrix}$ is

   $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and so the three vectors are

   linearly independent.

(b) The Hermite normal form of $\begin{bmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \\ 3 & 1 & 1 \end{bmatrix}$ is

$\begin{bmatrix} 1 & 0 & \frac{1}{4} \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}$, and so the first two vectors are linearly independent, and

$$(1, 1, 1) = \tfrac{1}{4}(1, 2, 3) + \tfrac{1}{4}(3, 2, 1).$$

(c) The Hermite normal form of $\begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & -2 & 1 & \frac{1}{2} \\ 2 & 4 & 2 & 5 \end{bmatrix}$ is

$\begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}$, and so $(1, -1, 2)$ and $(0, 1, 2)$ are linearly independent, and the other two vectors satisfy

$$(2, -2, 4) = 2(1, -1, 2)$$
$$(1, \tfrac{1}{2}, 5) = (1, -1, 2) + \tfrac{3}{2}(0, 1, 2).$$

2. The Hermite normal form of $\begin{bmatrix} 2 & 1 & 0 & 0 & 1 \\ 3 & 2 & 1 & 0 & 3 \\ 4 & 3 & 1 & 1 & 6 \\ 5 & 4 & 0 & 1 & 8 \end{bmatrix}$

is $\begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$

so that

$$(1, 3, 6, 8) = -1(2, 3, 4, 5) + 3(1, 2, 3, 4) + (0, 0, 1, 1).$$

3.4.3 Summary of Section 3.4

In this section we defined the terms

linear problem	(page N63)	* * *
particular solution	(page N63)	* * *
general solution	(page N64)	* * *
associated homogeneous problem	(page N64)	* * *
augmented matrix	(page N64)	* * *
row-sum check	(page C36)	* * *

We introduced the notation

$[A, B]$	(page N64)

Theorem

(*7.1*, page N64)
The system of simultaneous linear equations $AX = B$ has a solution if and only if the rank of A is equal to the rank of the augmented matrix $[A, B]$. Whenever a solution exists, all solutions can be expressed in terms of $\nu = n - \rho$ independent parameters, where ρ is the rank of A. * * *

Techniques

1. Solve $AX = B$ by reducing $[A, B]$ to Hermite normal form. * * *
2. Determine linear relations (if any) between vectors by reducing the matrix of coordinates of those vectors to Hermite normal form. * * *
3. Use the row-sum check. * * *

3.5 SUMMARY OF THE UNIT

Before the complete list of definitions, theorems, techniques and notation, here is a review of the unit.

Given a basis $\{\alpha_1, \ldots, \alpha_n\}$ of a vector space U, any vector $\xi \in U$ can be expressed in terms of $\{\alpha_1, \ldots, \alpha_n\}$ by means of a one-column matrix X consisting of the components of ξ. If we change the basis to $\{\alpha'_1, \ldots, \alpha'_n\}$, we can describe the change by a *transition matrix* P whose columns are the components of the new basis elements in terms of the old. The new one-column matrix X' for ξ is then related to X by the equation

$$X = PX'$$

If $\sigma: U \longrightarrow V$ is a linear transformation with a matrix A, and we change the bases in U and V by transition matrices P and Q respectively, then the new matrix for σ is $A' = Q^{-1}AP$. The simplest possible form for A' which we can obtain in this way, has $a'_{11} = a'_{22} = \cdots = a'_{\rho\rho} = 1$ (where ρ is the rank of σ), and all other matrix entries zero. If we are only allowed basis changes in the codomain (i.e. if we can only multiply A on the left by non-singular matrices), then the simplest possible form to which we can bring A is known as the *Hermite normal form*; when suitably defined, this form is unique.

The codomain basis $\{\beta_1, \ldots, \beta_m\}$ that achieves Hermite normal form can be characterized as follows. Look through the images of the domain basis, and apply the following rule. If $\sigma(\alpha_i)$ is linearly independent of $\sigma(\alpha_1), \ldots, \sigma(\alpha_{i-1})$, take $\sigma(\alpha_i)$ as the next basis element of the codomain; otherwise, reject it. When all the $\sigma(\alpha_i)$s have been exhausted, complete the basis of V arbitrarily.

A matrix A is in Hermite normal form if and only if it obeys the following four conditions (compressed into three on pages N54–55). Suppose A is $m \times n$.

(i) There is an integer ρ, $1 \leqslant \rho \leqslant m$, such that the first ρ rows of A are not zero rows and the last $m - \rho$ are.
(ii) Each of these first ρ rows has a 1 as its first non-zero entry.
(iii) This 1 is in each case the only non-zero entry in its particular column.
(iv) If this 1 is in the k_ith place in the ith row ($i = 1, \ldots, \rho$), then $k_1 < k_2 < \cdots < k_\rho$.

If A satisfies all four of the above conditions, then its rank is ρ. In practice, the way to get a matrix A into Hermite normal form is to apply elementary row operations repeatedly to it until it is in the required form. It is wise to apply the row operations in an orderly fashion, first of all getting the requisite zeros in the first column, then proceeding to the next column, and so on. When doing row operations by hand, the row sum check should always be used to guard against blunders.

An elementary row operation can be performed on A by left-multiplying A by the matrix which is obtained by applying the same row operation to the appropriate unit matrix. By reducing the augmented matrix $[A, I]$ to HNF, one obtains $[A', Q^{-1}]$, where A' is the HNF of A and Q^{-1} is the matrix by which A is left-multiplied to obtain A'. In particular, if A is non-singular, one obtains its inverse, A^{-1}, by this means.

The application of row operations does not change the linear relations that exist between the columns of a matrix. Thus, to find linear relations among a set of vectors, write these vectors as the columns of a matrix and reduce the matrix to Hermite normal form.

The solution set of a system of simultaneous linear equations is easy to find once the set has been converted to an equivalent set of equations by reducing the augmented matrix of the system to Hermite normal form.

As a slight digression from the main theme, we looked at the transpose of a matrix, in which the columns become rows and vice versa, and used this concept to prove that the row and column ranks of any matrix are equal.

Definitions

The terms defined in this unit and page references to their definitions are given below.

matrix of transition	(page N50)	* * *
similar	(page N52)	* * *
standard basis for R^n	(page C20)	* * *
Hermite normal form	(page N55)	* * *
transpose of a matrix	(page N55)	* * *
elementary row operation	(page N57)	* * *
elementary matrix	(page N58)	* * *
linear problem	(page N63)	* * *
particular solution	(page N63)	* * *
general solution	(page N64)	* * *
associated homogeneous problem	(page N64)	* * *
augmented matrix	(page N64)	* * *
row-sum check	(page C36)	* * *

Theorems

We list the theorems discussed in this unit. References to the statements of the theorems in N are also given.

1. (*4.1*, page N52)
   If A is any $m \times n$ matrix of rank ρ, there exist a non-singular $n \times n$ matrix P and a non-singular $m \times m$ matrix Q such that $A' = Q^{-1}AP$ has the first ρ elements of the main diagonal equal to 1, and all other elements equal to zero. * *

2. (*5.1*, page N54)
   Given any $m \times n$ matrix A of rank ρ, there exists a non-singular $m \times m$ matrix Q such that $A' = Q^{-1}A$ has the following form: * * *

   (1) There is at least one non-zero element in each of the first ρ rows of A' and the elements in all remaining rows are zero.
   (2) The first non-zero element appearing in row i $(i \leqslant \rho)$ is a 1 appearing in column k_i, where $k_1 < k_2 < \cdots < k_\rho$.
   (3) In column k_i the only non-zero element is the 1 in row i.

   The form A' is uniquely determined by A.

3. (***Proposition 5.2***, page N55)
   The number of linearly independent rows in a matrix is equal to the number of linearly independent columns. * * *

4. (*6.1*, page N58)
   Any non-singular matrix A can be written as a product of elementary matrices. * * *

5. (*7.1*, page N64)
   The system of simultaneous linear equations $AX = B$ has a solution if and only if the rank of A is equal to the rank of the augmented matrix $[A, B]$. Whenever a solution exists, all solutions can be expressed in terms of $v = n - \rho$ independent parameters, where ρ is the rank of A. * * *

Techniques

1. Determine the matrix of transition for a given change of basis. * * *
2. Determine the coordinates of a vector with respect to a new basis when given the coordinates of that vector with respect to an old basis and the appropriate matrix of transition. * * *
3. Determine the matrix of a linear transformation, σ, from U to V with respect to bases A' in U and B' in V, given the matrix representing σ with respect to bases A in U and B in V, and given the form of the change of bases A to A' in U and B to B' in V. * * *
4. Recognize when a matrix is in Hermite normal form and hence determine the rank of such a matrix. * * *
5. Given the matrix, A, representing a linear transformation $\sigma\colon U \longrightarrow V$ with respect to specified bases in U and V, determine the Hermite normal form of A by selecting a linearly independent subset of $\{\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\}$, $\{\alpha_1, \ldots, \alpha_n\}$ being the basis in U. * *
6. Use the properties of the transpose of a matrix in matrix manipulations. * * *
7. Use row operations to reduce a matrix to Hermite normal form. * * *
8. Use row operations to find the inverse of a matrix. * * *
9. Solve $AX = B$ by reducing $[A, B]$ to Hermite normal form. * * *
10. Determine linear relations (if any) between vectors by reducing the matrix of coordinates of those vectors to Hermite normal form. * * *
11. Use the row-sum check. * * *

Notation

U_k	(page N54)
$k_1, k_2, \ldots, k_p$	(page N54)
A^T	(page N55)
$E_2(c), E_{31}(k), E_{12}$, etc.	(page N57–58)
$[A, B]$	(page N64)

3.6 SELF-ASSESSMENT

Self-assessment Test

This Self-assessment Test is designed to help you test quickly your understanding of the unit. It can also be used, together with the summary of the unit for revision. The answers to these questions will be found on the next non-facing page. We suggest you complete the whole test before looking at the answers.

1. (i) Let $A = \{\alpha_1, \alpha_2\}$, $A' = \{\alpha'_1, \alpha'_2\}$ be bases in some vector space U. If

$$\alpha'_1 = 2\alpha_1$$

and $\alpha'_2 = \alpha_1 - 3\alpha_2$,

what is the matrix of transition from A to A'?

(ii) Let $B = \{\beta_1, \beta_2, \beta_3\}$, $B' = \{\beta'_1, \beta'_2, \beta'_3\}$ be bases in some other vector space V. If

$$\beta'_1 = \beta_1 - \beta_2 + 2\beta_3$$
$$\beta'_2 = -2\beta_2 + \beta_3$$
$$\beta'_3 = -3\beta_1 - \beta_3,$$

what is the matrix of transition from B to B'?

(iii) Let $\sigma: U \longrightarrow V$, such that

$$\sigma(\alpha_1) = \beta'_1 - 2\beta'_3$$
$$\sigma(\alpha_2) = -3\beta'_1 + \beta'_2$$

What is the matrix representing σ with respect to A and B'?

(iv) Using your answers to parts (i), (ii) and (iii), determine the matrix representing σ with respect to A' and B.

2. Find the Hermite normal form of

$$\begin{bmatrix} 2 & 4 & 1 & 10 \\ 1 & 2 & 1 & 7 \\ 0 & 0 & 1 & 4 \end{bmatrix}.$$

Use the row-sum check.

3. Find a linearly independent subset of

$$\{(2, 1, 0), (4, 2, 0), (1, 1, 1), (10, 7, 4)\}$$

and express each of those elements that are not in this subset, as linear combinations of those that are.

4. Find the inverse of $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{bmatrix}$ by determining its Hermite normal form.

5. The vectors $\beta_1, \beta_2, \beta_3, \beta_4$ in R^3 are written as the columns of a matrix B. β_1, β_2 and β_3 are linearly independent, and

$$\beta_4 = \beta_1 + \beta_2 - \beta_3$$

The second and third rows of the matrix are interchanged, the new columns being

$$\beta'_1, \beta'_2, \beta'_3, \beta'_4$$

(i) Are $\beta'_1, \beta'_2, \beta'_3$ linearly independent?

(ii) Which of the following is true?

(a) $\beta'_4 = \beta'_1 - \beta'_2 + \beta'_3$
(b) $\beta'_4 = \beta'_1 + \beta'_2 - \beta'_3$
(c) Neither (a) nor (b) holds

6. A is a 4×4 matrix of rank 1. The first column of A contains some non-zero elements. Write down the general form of the Hermite normal form of A subject to these conditions.

7. U and V are 3-dimensional vector spaces. With respect to bases $\{\alpha_1, \alpha_2, \alpha_3\}$ of U, $\{\beta_1, \beta_2, \beta_3\}$ of V, the matrix of a linear transformation $\sigma: U \longrightarrow V$ is

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}.$$

Write down the matrix of σ with respect to:

(i) $\{\alpha_3, \alpha_2, \alpha_1\}$ and $\{\beta_1, \beta_2, \beta_3\}$

(ii) $\{\alpha_1, \alpha_2, \alpha_3\}$ and $\{\beta_3, \beta_2, \beta_1\}$.

Solutions to Self-assessment Test

1. (i) $\begin{bmatrix} 2 & 1 \\ 0 & -3 \end{bmatrix}$ (ii) $\begin{bmatrix} 1 & 0 & -3 \\ -1 & -2 & 0 \\ 2 & 1 & -1 \end{bmatrix}$

The column entries are the coordinates of the new basis elements in terms of the old ones.

(iii) Since $\sigma(\alpha_1) = \beta'_1 + 0\beta'_2 - 2\beta'_3$
$\sigma(\alpha_2) = -3\beta'_1 + \beta'_2 + 0\beta'_3$

the matrix of σ with respect to A and B' is

$$\begin{bmatrix} 1 & -3 \\ 0 & 1 \\ -2 & 0 \end{bmatrix}$$

(iv) The matrix representing σ with respect to A' and B is $M' = Q^{-1}MP$, where
P is the matrix of transition from A to A',
M is the matrix representing σ with respect to A and B',
Q^{-1} is the inverse of the matrix of transition from B' to B; i.e. Q^{-1} is the matrix of transition from B to B'.

Thus

$$P = \begin{bmatrix} 2 & 1 \\ 0 & -3 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & -3 \\ 0 & 1 \\ -2 & 0 \end{bmatrix}$$

and

$$Q^{-1} = \begin{bmatrix} 1 & 0 & -3 \\ -1 & -2 & 0 \\ 2 & 1 & -1 \end{bmatrix}$$

and hence

$$M' = \begin{bmatrix} 1 & 0 & -3 \\ -1 & -2 & 0 \\ 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 0 & 1 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & -3 \end{bmatrix}$$

$$= \begin{bmatrix} 14 & 16 \\ -2 & -4 \\ 8 & 19 \end{bmatrix}$$

2. The Hermite normal form of

$$\begin{bmatrix} 2 & 4 & 1 & 10 \\ 1 & 2 & 1 & 7 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$

is

$$\begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

row sums

For, row 1 becomes $\frac{1}{2}$ row 1 gives $\left[\begin{array}{cccc|c} 1 & 2 & \frac{1}{2} & 5 & 8\frac{1}{2} \\ 1 & 2 & 1 & 7 & 11 \\ 0 & 0 & 1 & 4 & 5 \end{array}\right]$

row 2 becomes row 2 − row 1 gives $\left[\begin{array}{cccc|c} 1 & 2 & \frac{1}{2} & 5 & 8\frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 2 & 2\frac{1}{2} \\ 0 & 0 & 1 & 4 & 5 \end{array}\right]$

row 1 becomes row 1 − row 2, row 3 becomes row 3 − 2 row 2 give

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & 3 & 6 \\ 0 & 0 & \frac{1}{2} & 2 & 2\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

and finally row 2 becomes 2 row 2 gives

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & 3 & 6 \\ 0 & 0 & 1 & 4 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

3. The vectors in question are just the columns of the matrix in question 2. The solution to question 2 gives the required answer. Thus (2, 1, 0) and (1, 1, 1) are linearly independent and

$$(4, 2, 0) = 2(2, 1, 0)$$
$$(10, 7, 4) = 3(2, 1, 0) + 4(1, 1, 1).$$

4. The inverse of $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{bmatrix}$ is $\begin{bmatrix} 3 & -1 & -2 \\ -1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix}$.

For, we can find the Hermite normal form of

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 & 1 \end{bmatrix}$$

as follows.

row 1 becomes row 1 − row 2

$$\begin{bmatrix} 1 & 0 & 2 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 & 1 \end{bmatrix}$$

row 3 becomes row 3 − row 1

$$\begin{bmatrix} 1 & 0 & 2 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 1 \end{bmatrix}$$

row 3 becomes row 3 − row 2

$$\begin{bmatrix} 1 & 0 & 2 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}$$

Finally, row 1 becomes row 1 − 2 row 3, and
row 2 becomes row 2 + row 3

$$\begin{bmatrix} 1 & 0 & 0 & 3 & -1 & -2 \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}$$

5. The linear relations among $\beta'_1, \beta'_2, \beta'_3, \beta'_4$ are just the same as those among $\beta_1, \beta_2, \beta_3, \beta_4$. Thus

(i) $\beta'_1, \beta'_2, \beta'_3$ *are* linearly independent.
(ii) $\beta'_4 = \beta'_1 + \beta'_2 - \beta'_3$.

6. Since the first column of A contains some non-zero elements, the first column of the Hermite normal form A', must be

$$\begin{matrix} 1 \\ 0 \\ 0 \\ 0 \end{matrix}$$

But A' is of rank 1; hence the other three columns of A' must be dependent on this column. So the general form of A' is

$$\begin{bmatrix} 1 & a & b & c \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

7. (i) Let $A = \{\alpha_1, \alpha_2, \alpha_3\}$
$A' = \{\alpha_3, \alpha_2, \alpha_1\}$
and $B = \{\beta_1, \beta_2, \beta_3\}$.

Then the matrix of transition from A to A' is

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

and hence the matrix representing σ with respect to A' and B is

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 7 & 4 & 1 \\ 8 & 5 & 2 \\ 9 & 6 & 3 \end{bmatrix}.$$

(Note that $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ is the elementary matrix which interchanges columns 1 and 3.)

(ii) Similarly, if $B' = \{\beta_3, \beta_2, \beta_1\}$, the matrix of transition from B to B' is

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

and the matrix representing σ with respect to A and B' is

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 6 & 9 \\ 2 & 5 & 8 \\ 1 & 4 & 7 \end{bmatrix}$$

(Again, notice the action of the elementary matrix in interchanging rows 1 and 3.)